OLLSCOIL NA GAILLIMHE
UNIVERSITY OF GALWAY

MS5115- Business Analytics Major Project

Topic: Sports Analysis of football (At a club level)

Project Supervisor
Prof. Lorraine Morgan

Team members
Abhishek Kumar (Project Manager)
Harsh Mahesh Kulkarni
Mohammad Shavez Khan
Rajat Rajendra Kelkar
Sajal Kothari

Declaration of Originality

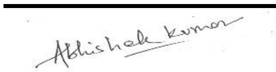
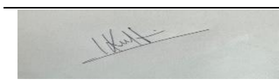
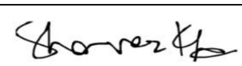
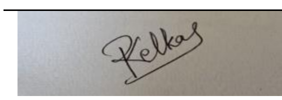
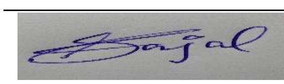
Project Details

Module Code: **MS5115** Assignment Title: **M.Sc. Business Analytics - Major Project**

Group Members: Fourteen (14)

Student ID	Student Name	Contact Details (Email, Telephone)
22221703	Abhishek Kumar	a.kumar10@nuigalway.ie, +353899888595
22221739	Harsh Mahesh Kulkarni	h.kulkarni1@nuigalway.ie, +353899411163
22221713	Mohammad Shavez Khan	M.khan19@nuigalway.ie, +353831244101
22225434	Rajat Rajendra Kelkar	r.kelkar1@nuigalway.ie, +91 96195 69988
22220073	Sajal Kothari	s.kothari1@nuigalway.ie, +353857304440

I/We hereby declare that this project is my/our own original work. I/We have read the University *Code of Practice for Dealing with Plagiarism*[*] and am/are aware that the possible penalties for plagiarism include expulsion from the University.

Signature	Date
	06/08/2023
	06/08/2023
	06/08/2023
	06/08/2023
	06/08/2023

[*] http://www.nuigalway.ie/plagiarism

Team profile

Abhishek Kumar (Project Manager)

Professional with a 7 years of experience, working in the research industry. Skilled in Statistics, Healthcare Analytics, Business Analytics, Market Research, and Project Management. In software, I have excellent competency in R, Python, Tableau, SQL, SPSS and MS excel. Qualified with MBA focused on Rural Marketing Management from Xavier Institute of Social Service (XISS), India and MSc. Business Analytics from University of Galway, Ireland.

Harsh Mahesh Kulkarni

Software Development Engineer with 4 years of Product development experience in BFSI and Integration domain. I specialise in designing and developing backend systems of a Product with scalable solutions using Cloud. Being a self-taught programmer, I have always relied on my curiosity to identify potential opportunities and learning new skills.

Mohammad Shavez Khan

A dedicated and detail-oriented student pursuing my MSc in Business Analytics from NUIG with a strong foundation in the field of network engineering, backed by over 3 years of industry experience

Rajat Kelkar

Results-driven Business Analyst with a strong track record of delivering impactful solutions and driving business growth. Skilled in data analysis, requirements gathering, and project management, I have collaborated with cross-functional teams to optimize processes and enhance organizational performance. Currently I am pursuing my Masters in Business Analytics from University of Galway.

Sajal Kothari

Business Operations Analyst with 3.5 years of experience in the Insurance and Finance Domain. My expertise lies in conducting intricate analysis and utilizing data-driven decision-making to improve finance operations and drive business growth. My focus has enabled me to identify opportunities for improvement and implement effective solutions.

Acknowledgement

We would like to express our sincere gratitude and appreciation to all those who have contributed to the completion of this project report. Without their support, guidance, and encouragement, this endeavor would not have been possible.

First and foremost, we are deeply thankful to our project supervisor Prof. Lorraine Morgan, for their invaluable guidance, expertise, and unwavering support throughout the duration of this project. Their insightful feedback and constructive suggestions have been instrumental in shaping the direction and quality of this report.

We would also like to extend our gratitude to the faculty members of the department at the university of Galway for imparting their knowledge and providing a conducive learning environment. Their passion for teaching and dedication to our academic growth have greatly enriched my understanding of the subject matter. Their cooperation and involvement have been vital in gathering the necessary information and ensuring the reliability of our findings.

We would like to acknowledge the assistance and support received from our friends and classmates. Their encouragement, discussions, and shared experiences have been immensely valuable in enhancing my understanding and motivation during the project.

Furthermore, we are indebted to the authors, researchers, and scholars whose work has served as a foundation for this project. Their contributions have laid the groundwork for our research and provided valuable insights and perspectives in the field of study.

Lastly, we would like to express our heartfelt appreciation to our families for their constant love, encouragement, and understanding. Their unwavering support and belief in my abilities have been a source of strength throughout this journey.

In conclusion, we extend our gratitude to everyone who has played a role, big or small, in the completion of this project report. Your contributions have been instrumental in its success, and we are truly grateful for your support.

Thank you.

Team-14

Executive summary

Sports analytics has progressed as a crucial tool in the world of sports, delivering data-driven insights to teams, individuals, and organisations in order to achieve a competitive improvement. With technological improvements and the availability of large amounts of data, the importance of using analytics has been recognised. Sports analytics entails the collection, analysis, and interpretation of numerous sorts of data, such as player performance statistics, team plans, game circumstances, and spectator behaviour. Sports analytics offers performance optimisation, informed decision-making, and increased fan engagement by leveraging these insights.

One of the most conventional applications of sports analytics is to improve team and athlete performance. Teams can detect patterns, assess strengths and weaknesses, and make informed judgements on tactics, training techniques, and game strategies by analysing player performance data. For example, analysing player tracking data allows teams to identify areas for growth, such as shooting accuracy or field decision-making. Coaches and trainers can then create customised training programmes to improve team performance and overall success.

Sports analytics is key in the evaluation and recruiting of players. Teams can find talented individuals, appraise their talents, and assess their prospective impact on the team by analysing player performance indicators. Various data points, including scoring efficiency and defensive measures, can help make informed decisions on signings, transfers, and draught picks. This data-driven strategy assists teams in making objective and informed roster and player investment decisions.

The use of sports analytics in football is the result of interdisciplinary partnership between data scientists, researchers, and football experts. Their combined efforts have paved the path for advanced approaches, data-driven insights, and practical applications in the field of sports analytics.

As the field develops, new challenges and opportunities emerge. Ethical concerns, data privacy, and the responsible use of analytics in decision-making continue to be hot topics. Nonetheless, advances in sports analytics continue to reshape how football is understood, played, and managed. The same we are trying to develop in this project with Liverpool premier league data.

We have divided the study area and the associated analysis into the four major categories, i.e.,

- Financial analysis,
- Team Management & Player Performance,
- Club Management and Administration,
- Predictive analytics in sports for team and detailed player's analytics.

There were numerous web pages with information and statistics about football matches and events. The information pertains to both teams and players. Some data was accessed and obtained manually, especially when it was simple. Some of them, however, were scraped off the internet using various scraping technologies. Data of European football league (Premier league) was taken. Finally, a database was created and utilised in the studies. The database was collected from a popular manager simulation game and contains data from thousands of participants. It displays player ratings for various football skills. Domain specialists rate the players. Following the data collection procedure, there was a vast database that needed to be organised. The database was divided into various csv files based on what data was required for each experiment.

The analysis resulted in various outcomes that can be looked upon to improve the performance of the squad as a whole. The data collected and analyzed was for the last 20 years and the pattern on points and the half-time scores and the end result were studied over the years. Exploratory data analysis was done to understand the team dynamics and study the goals scored, conceded and points won in the previous seasons. The predictive model that was built to find the probability of the win on the basis of halftime results and the opposition and also the venue being played on shows that it can be used to motivate the team to score more goals and go for a win. In order to increase the model accuracy, we can run various other models and increase the number of input parameters and take into consideration the time between the games, whether they are played on the weekend or the weekdays and how they impact the result. Overall, the exploratory analysis that was done helped to understand the past results and the team performed within various matrices.

Liverpool could look to sign new players in the transfer window to address their weaknesses. They could target a new centre-back or full-back to improve their defense, or a striker or attacking midfielder to provide more firepower in the final third. Liverpool could focus on improving their overall team shape and positioning to ensure that they are more solid defensively and more dangerous going forward. This could involve working on their defensive transitions and pressing, as well as their attacking movements and combination play. Liverpool could focus on improving their set-piece play to generate more chances and goals. This could involve developing new set-piece routines or practicing existing ones to make them more effective. Also, Liverpool could work on improving their mental toughness and resilience, particularly in tight matches. This could involve working on their concentration, focus, and confidence to ensure that they are able to see out games and grind out results when needed.

First 45 min and conversion rate should be checked. Encourage the team to take the lead at halftime, as seen from the data that the results remain the same in the majority of cases. The team needs to improve the conversion rate from set pieces, hiring a dedicated set piece coach can be done. Also, the team needs to improve the away record if we have to challenge for the title.

In order for football clubs to not only survive but also succeed in the fiercely competitive sports industry, it is crucial for them to find a balance between their financial stability and team performance. By following the provided guidance, clubs can make better-informed decisions, ultimately leading to a more prosperous and flourishing organization.

Table of Contents

List of Figures

List of tables

1. Introduction

Sports analytics has progressed as a crucial tool in the world of sports, delivering data-driven insights to teams, individuals, and organisations in order to achieve a competitive improvement. With technological improvements and the availability of large amounts of data, the importance of using analytics has been recognised. Sports analytics entails the collection, analysis, and interpretation of numerous sorts of data, such as player performance statistics, team plans, game circumstances, and spectator behaviour. Sports analytics offers performance optimisation, informed decision-making, and increased fan engagement by leveraging these insights.

One of the most conventional applications of sports analytics is to improve team and athlete performance. Teams can detect patterns, assess strengths and weaknesses, and make informed judgements on tactics, training techniques, and game strategies by analyzing player performance data. For example, analyzing player tracking data allows teams to identify areas for growth, such as shooting accuracy or field decision-making. Coaches and trainers can then create customized training programs to improve team performance and overall success.

Sports analytics is key in the evaluation and recruiting of players. Teams can find talented individuals, appraise their talents, and assess their prospective impact on the team by analyzing player performance indicators. Various data points, including scoring efficiency and defensive measures, can help make informed decisions on signings, transfers, and draught picks. This data-driven strategy assists teams in making objective and informed roster and player investment decisions.

Analytics betters game tactics and decision making by providing teams with useful information about their opponents' playing styles, strengths, and weaknesses. Analyzing historical data allows for the identification of opponent tendencies, patterns, and distinct methods used by various teams and players. Coaches can then create game plans and make in-game modifications to provide their team with a competitive advantage. Real-time analytics aid decision making by providing quick feedback on player performance, opponent behavior, and ideal substitution patterns throughout games.

Sports analytics stretches beyond on-field uses to understanding spectator behavior and driving corporate decisions. Organizations can customize marketing efforts, improve fan experiences, and make educated business decisions by analyzing data linked to ticket sales, goods purchases, social media interactions, and fan sentiment. Teams can use this data-driven strategy to discover target groups, personalize fan experiences, and optimize revenue sources through ticket sales, sponsorships, and collaborations.

Sports analytics has become influential in sports, allowing teams, athletes, and organizations to use data-driven insights to upgrade performance, player evaluation and recruiting, injury prevention and management, game strategy, and fan engagement. Sports analytics gives a competitive advantage in an industry where even minor advantages count by leveraging the power of data analysis and statistical modelling. The relevance of sports analytics will only expand as technology advances, giving new opportunities for teams and organizations to improve their understanding of their sport and maximize their chances of success both on and off the pitch.

2. Background

In recent years, sports analytics has gained tremendous traction, giving useful insights and data-driven solutions to optimize performance, enhance decision-making, and gain a viable edge in numerous sports disciplines. Within this field, football has emerged as a significant area of focus due to its global popularity, complex dynamics, and the wealth of data generated during matches.

Football, or soccer, is a team sport known for its fast-paced nature, tactical gameplay, and intricate interplay among players, teams, and tactical formations. Traditionally, evaluating player performance and game strategies relied on subjective observations and limited statistical metrics. However, developments in technology and the availability of rich data sources, such as player tracking systems, match statistics, and video analysis, have opened up new possibilities for comprehensive analysis in football.

Sports analytics in football comprises the collecting, analysis, and interpretation of varied data sources to obtain insights into player execution, team dynamic range, and strategic decision-making. Through statistical modeling approaches, machine learning algorithms, and data visualization tools, academics and practitioners can uncover designs, identify trends, and extract useful insights from enormous football-related data.

The use of sports analytics in football covers a wide range of aspects of the game. For example, data-driven initiatives have altered player evaluation and recruitment processes. Metrics like Expected Goals (xG), Pass Completion Rate, and Defensive Actions per 90 minutes offer more comprehensive player profiles, enabling clubs to make informed decisions when scouting and acquiring new talent (Koehn et al., 2019). Analytics also contribute to injury prevention and management by analyzing workload data, monitoring player fatigue levels, and identifying potential injury risks (Buchheit, 2014).

Additionally, sports analytics have had a substantial effect on football game strategy and decision-making. Trainers can use advanced metrics like Expected Possession Value (EPV), Pass Network Analysis, and Space Control models to identify effective playing patterns, optimize player positioning, and devise game plans to exploit opponents' weaknesses (Decroos et al., 2019; Fernandez-Navarro et al., 2019). Coaches can make on-the-fly adjustments and substitutions based on player performance indicators and opponent behavior using real-time analytics during matches (Duch et al., 2019).

The use of sports analytics in football is the result of interdisciplinary partnership between data scientists, researchers, and football experts. Their combined efforts have paved the path for advanced approaches, data-driven insights, and practical applications in the field of sports analytics.

As the field develops, new challenges and opportunities emerge. Ethical concerns, data privacy, and the responsible use of analytics in decision-making continue to be hot topics. Nonetheless, advances in sports analytics continue to reshape how football is understood, played, and managed. The same we are try to develop in this project with Liverpool premier league data.

3. Analytical methodology

We have divided the study area and the associated analysis into the four major categories, i.e.,

- Financial analysis,
- Team Management & Player Performance,

- Club Management and Administration,
- Predictive analytics in sports for team and detailed player's analytics.

First, relevant data about football clubs was searched. There were numerous web pages with information and statistics about football matches and events. The information pertains to both teams and players. Some data was accessed and obtained manually, especially when it was simple. Some of them, however, were scraped off the internet using various scraping technologies. Data of European football league (Premier league) was taken. Finally, a database was created and utilised in the studies. The database was collected from a popular manager simulation game and contains data from thousands of participants. It displays player ratings for various football skills. Domain specialists rate the players. Following the data collection procedure, there was a vast database that needed to be organised. The database was divided into various csv files based on what data was required for each experiment. The csv files were then transferred to Jupiter, R, and other tools for data processing. Naturally, the data had to be pre-processed first. They were examined for null values, duplicates, noise, and so forth. To clean the data and develop the models, Python was utilised. The data was then transformed and reduced to retain only the relevant features for each classification or regression.

3.1 Financial analysis

Football club financial analysis entails evaluating their financial condition and performance, as well as investigating their transactions, financial assets, and strategic plan to identify areas for improvement. This can include analyzing their financial performance and identifying any areas of risk by reviewing their balance sheets, income statements, and cash flow statements. Furthermore, the analysis can consider the club's overall strategy, including marketing initiatives, media exposure, and overall operations.

3.1.1 Objective and business question

The goal is to use data-driven insights and predictions to assist these stakeholders in making more informed decisions about player transfers, scouting, and other related activities. There are several reasons why developing a recommender system could be advantageous:

- Increased efficiency: By simplifying the data search process, recommender systems can rapidly determine the most suitable transfer targets for clubs and players, saving both time and money.
- Better decision making: Recommender systems can assist clubs and players in making more informed choices about player transfers and scouting, lowering the risk of poor decisions and continuing to increase the likelihood of success.
- Increased revenue: A recommender system can help clubs and players increase revenue and performance by improving transfer and scouting decisions, which leads to increased success and sales.

3.1.2 Analysis plan

The best machine learning algorithm for building a recommender system in soccer sports analytics is determined by the specific problem at hand as well as the features of the data. Among the most used techniques which we are going to use for developing recommender systems are:

- Expense analysis: We examine the athletic department's costs and pinpoint the main cost factors. We categorize the costs and compare them to past expenditure patterns or industry norms. This study might point out places where money is being wasted or where money might be saved.
- Revenue analysis: The main sources of funding for the athletic department can be determined by analyzing the money it generates. We can segment the revenue into categories and assess the performance against standard metrics for the sector or previous patterns. This research can be used to pinpoint possible revenue sources or growth regions.
- Profitability analysis: We examine the athletic department's net income and contrast it with standard operating procedures or previous patterns. This examination helps determine the department's financial stability and capacity to continue operating over time.
- Benchmarking analysis: The athletic department's financial performance might be compared to that of other divisions or conferences. To make strategic decisions, this study can assist identify areas of strength or weakness in comparison to peers.
- Clustering Model: Utilize K Means revenue clustering to classify NCAA institutions according to their sources of income. This can offer insightful information about the various revenue models that institutions use and how they stack up against others in their cluster. We are able to display the clustering findings on a scatter plot by creating revenue categories, one-hot encoding the data, standardizing it, and lowering dimensionality using PCA. By doing so, it may be possible to spot patterns and connections that are hidden in the raw data. In conclusion, revenue clustering can be a beneficial tool for comprehending the financial performance of NCAA universities and guiding strategic choice-making.
- Time-Series Model: Apply several time series analytic techniques to the NCAA Profit and Losses dataset to demonstrate their use. Understanding the trends in total revenue over time and making revenue projections are the objectives of this analysis.

3.2 Team Management & Player Performance analysis

This category involves analysis of various aspects of team management and player performance including Team formation, Player stats, Sustainability & potential future of the player. We aim to verify whether these factors does/does not play an important role in club expansion, both financially & PR-wise. This would also help in gaining insights of the Team as a whole and factors that need to be altered to maximise overall club profit.

3.2.1 Objective and business question

Overall objective is to provide insights on the data to better recommend advice for the sake of the Player selection, Team performance and Players who do/don’t fit in the team.

3.2.2 Analysis plan

. Following are the specifics of the analysis that would be taken into consideration:

- Formation success rate: The success rate of team formation can be evaluated, which can indicate the effectiveness of the coach in selecting the right players and formation for the team.

- Average Player ratings: The Player ratings could help evaluate the position and the role of the player in the team. As a substitute, rotate in positions with highest percentage of holding capacity of the ball, etc.
- Average minutes played: This will help in getting insights on how many players are given less minutes and the accordingly decide their future.
- SWOT Aspects of team: A SWOT analysis of the team can also be conducted to identify its strengths, weaknesses, opportunities, and threats. This can help in developing strategies to improve the team's performance.
- Average goal conceded by team: This stat can assist in affective selection of defender's rotation eventually leading to a fit team of playing 11's with solid defense.
- Popularity: Club stats such as ratings for merchandise upscale and popularity can indicate their marketability, which can impact the team's revenue.
- Player Performance for evaluation of loan and sell: Evaluating each player's performance can help in making decisions about their loan or sale.

3.3 Club Management and Administration analytics

This section deals with analysis regarding the season ticket holders, C-suite expenses, scouting budget and player search, travel, and the marketing strategy and dependency on sponsorship. The above-mentioned factors play an important role in managing a club and directly/indirectly affect the overall structure of the club. The term sporting project comes in picture here, where we try to attract the promising young talent out there by providing them with the best launch pad for our club. The transfer market policies will also be studied in this section.

3.3.1 Research Problem/objective/business question

Player scouting and signing new players in the transfer market based on the performances over the years is the area that is in major focus here. Along with that marketing strategies such as defining goals, market research, resource allocation etc. to enhance the reach of the club and get better exposure worldwide will also be looked upon. Also, the objectives of the analysis are set. The amount of money that the club is spending on new player search is ripping the dividends in terms of player performance is also taken into consideration. Coming to the marketing part, we have tried to analyze the merchandise sale, the steps to improve the marketing strategies and its pattern over the years. Season ticket holders and the facilities that they get are also explored. Sponsorship can play a role in building the club, which also will be discussed. The number of sponsors a football club has varies greatly depending on various factors such as the club's size and popularity, geographic location, and economic situation. Few of the football clubs have just a handful of sponsors, while others have dozens or even hundreds. Also, the types of sponsors vary, with some clubs reliant mainly on corporate sponsors for economic support, while others have a mixture of corporate supporters and individual donors or supporters. The number and type of sponsors a football club has depends on many factors and can vary over time.

Our research focuses on statistics from the previous season and historical data. The analysis focuses on studying the available data and help in better decision making in years to come.

3.3.1 Analysis plan

Transfer market analysis/promotion. The features used for the experiment are the ones that were considered more relative to team performance. Those features can be divided into 4 categories. Players picked in the last 10 seasons and their on-field performance, Manager changes over the years, Platforms for promotion of brand.

Study of number of scouts in different countries and players coming from the youth system. A descriptive statistic will be used to describe the results.

Merchandise sale analysis: Pattern of merchandise sale over last 10 years and countries where most merch was sold. A descriptive statistic will be used to describe the results.

Sponsorship: Number of sponsors and how sponsors help in generating overall revenue. A descriptive statistic will be used to describe the results.

3.4 Predictive analytics in sports for team and detailed players analysis

Common Machine Learning applications in sports analytics relate to potential skill or market value evaluation, as well as team or player performance prediction. Its scope is long term team and player performance prediction. A reliable prediction of the final league table for certain leagues will be presented, using past data and advanced statistics. Other predictions for team performance include whether a team is going to have a better season than the last one.

3.4.1 Objective and business question

Long–term performance prediction for teams or individual players are fields requiring exploration. Not only coaches, but also sports agents and bookmakers are interested in how teams or players perform during a season compared to previous ones.

In this section, our research focuses on statistics from the previous season and historical data. The novelty of this research is that advanced metrics were used, such as goals conceded and results at home and away matches to predict next season performance before the season begins, not after matches have already been played and recorded. Additionally, attackers are usually graded higher than defenders, even if they are not always more influential in team strategy. So, regarding player evaluation, this research attempts to identify skills and features suitable, that make good defenders in the transfer market.

3.4.2 Analysis plan

Team Performance Prediction: The features used for the experiment are the ones that were considered more relative to team performance. Those features can be divided in two categories:

- Past data generated during the last 25 years. This mainly refers to performance indicators from previous seasons (e.g., team average points). We will use predictive modelling.
- Team statistical features from the season that has just ended (e.g., wins, shots, possession percentage and more).

Exploratory data analysis to understand the playing patterns.

- Goals conceded from corners over the last 25 seasons.
- League table position analysis
- Home and away record analysis

Apart from this we also tried to make a radial diagram for the players based on their performance. And a predictive model to predict the goal scoring ability based on different parameters like age. height, weight etc.

4. Findings of analysis

4.1 Financial Analysis

1st Experiment: Establish a list of expense categories before figuring out the overall costs and percentages for each. Then make a pie chart to show the breakdown of costs. Print the category-by-category expense breakdown in both dollars and percentages.

This information on athletic department costs can be used for analysis or presentation. The printed breakdown and pie chart give an understandable visual and numerical breakdown of the various spending categories. By doing so, it will be easier to spot areas where the department is spending money and make better resource allocation choices.

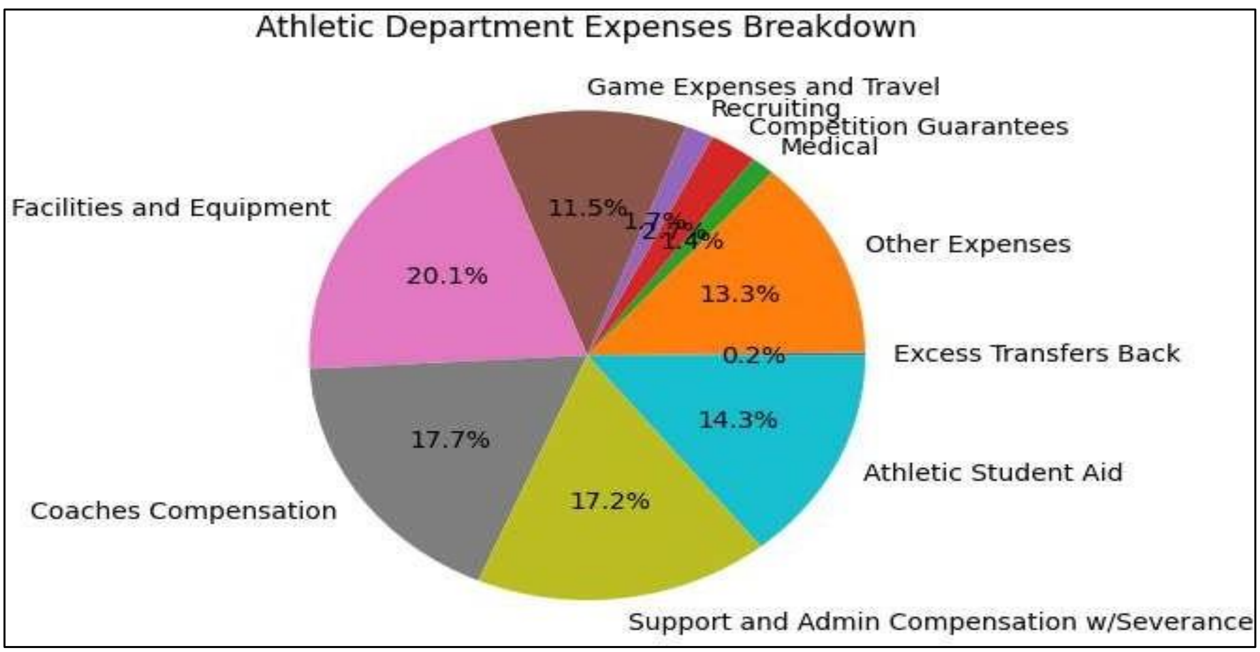


Figure 1: Athletic department expenses breakdown

The analysis shows that the major expense goes to facilities and equipment followed by coaches and support and admin.

2nd Experiment: Athletic directors, NCAA officials, and other stakeholders could find this analysis to be quite insightful. Each NCAA division and FBS conference's overall spending are shown graphically by the heatmap. Administrators and sports directors might find areas where they might be able to cut costs or better manage resources by comparing the total costs throughout the subdivisions and conferences. This approach can also be used to examine individual sporting programs' financial success and spot trends over time.

The application of data analysis methods to uncover information on the financial health of NCAA athletic programs. Administrators and stakeholders can manage sports programs effectively by allocating resources and considering the trends and factors that influence costs.

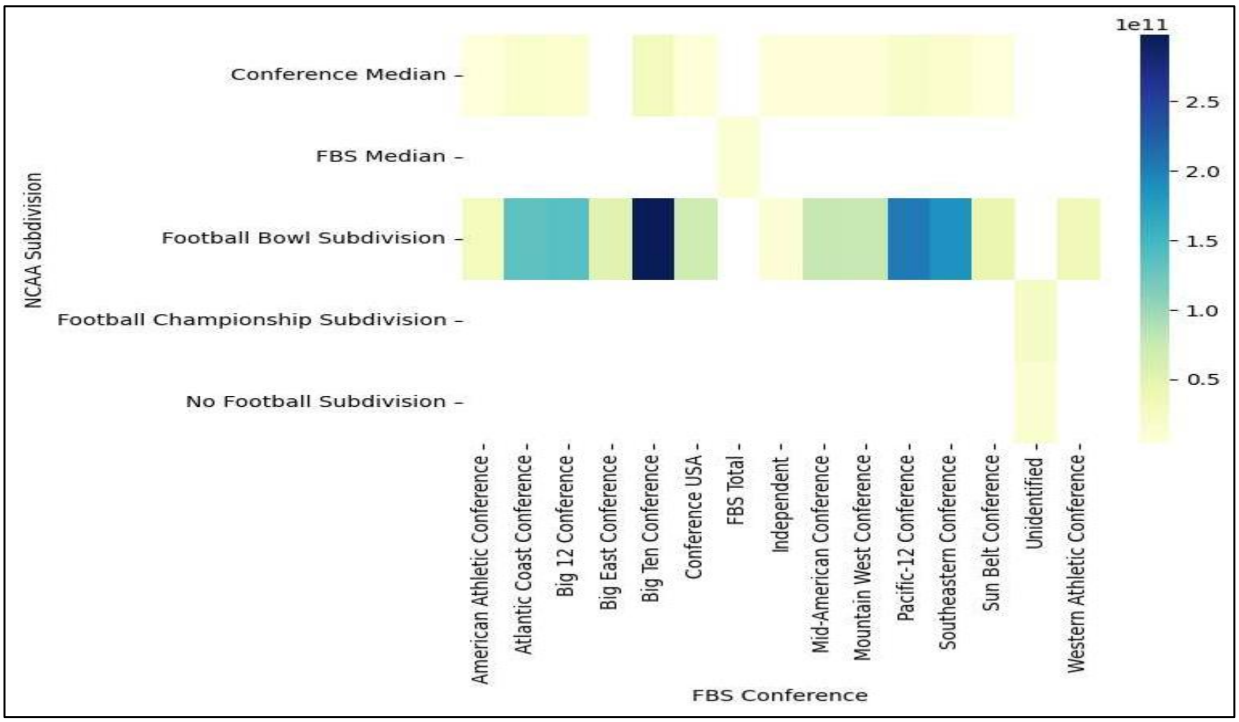


Figure 2: Overall spending are shown graphically by the heatmap.

3rd Experiment: An athletic department's revenue breakdown by category is shown in a pie chart. Having a clear grasp of where the revenue is coming from and how much each category adds to the overall revenue is one advantage of doing this. This can be helpful in identifying areas of strength or weakness in the department's revenue streams and in decision-making processes such as where to spend resources and funds.

Overall, any business, including an athletic department, would benefit from having a solid understanding of the revenue breakdown by category as it can aid in financial planning, budgeting, and resource allocation.

4th Experiment: To see the percentage of overall revenue earned by each revenue category, a scatter plot is made. The scatter plot makes it simple to see which revenue streams are most responsible for the athletic department's overall revenue.

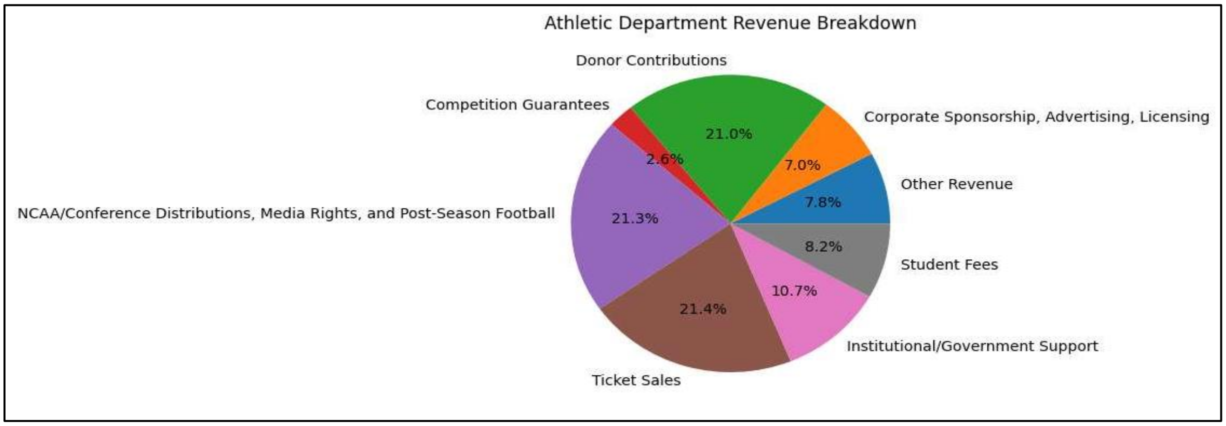


Figure 3: Athletic department revenue breakdown

The sports department can decide which revenue categories to concentrate on and potentially enhance their revenue streams by carefully examining the revenue breakdown. For instance, the sports department may decide to pursue tactics to enhance ticket sales revenue, such as providing promotions, lowering ticket prices, or enhancing the fan experience, if ticket sales are not making up a sizable portion of the total revenue. To improve the financial health of the clubs, the athletic department can make data-driven decisions by looking at the revenue breakdown.

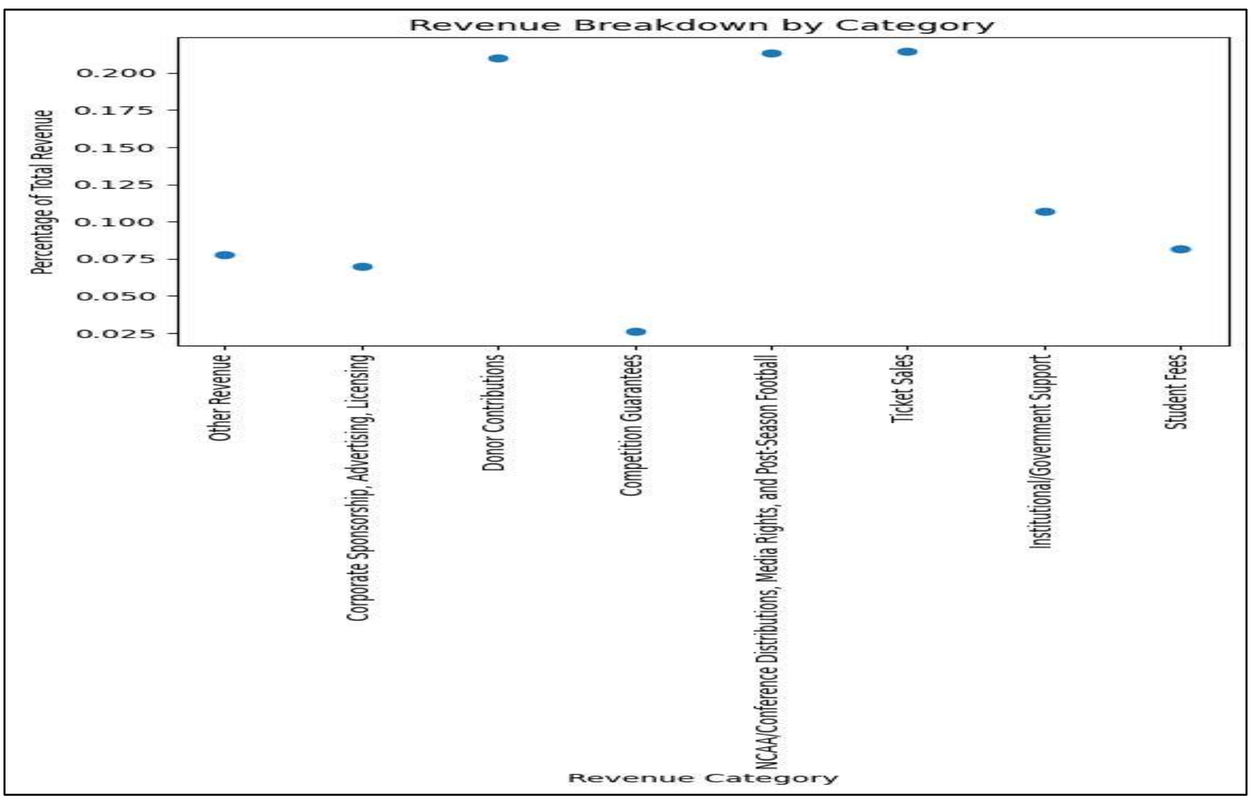


Figure 4: Revenue breakdown by category

5th Experiment: Analyze the profitability of a dataset that includes financial data for NCAA athletic departments. Create categories for revenue and outlays. The total revenue, cost, and profit are then determined for each FBS Conference in the dataset.

The first graphic displays the overall schools' profit and loss over time. The second plot displays each school's profitability, arranged from highest to lowest profit.

Finally, use the K Means technique to cluster the dataset according to revenue and spending categories. Then, make a scatter plot with points colored according to cluster membership, showing total earnings vs. total expenses for each institution.

In general, this research shows which clubs are the most lucrative as well as how different revenue and spending categories affect profitability.

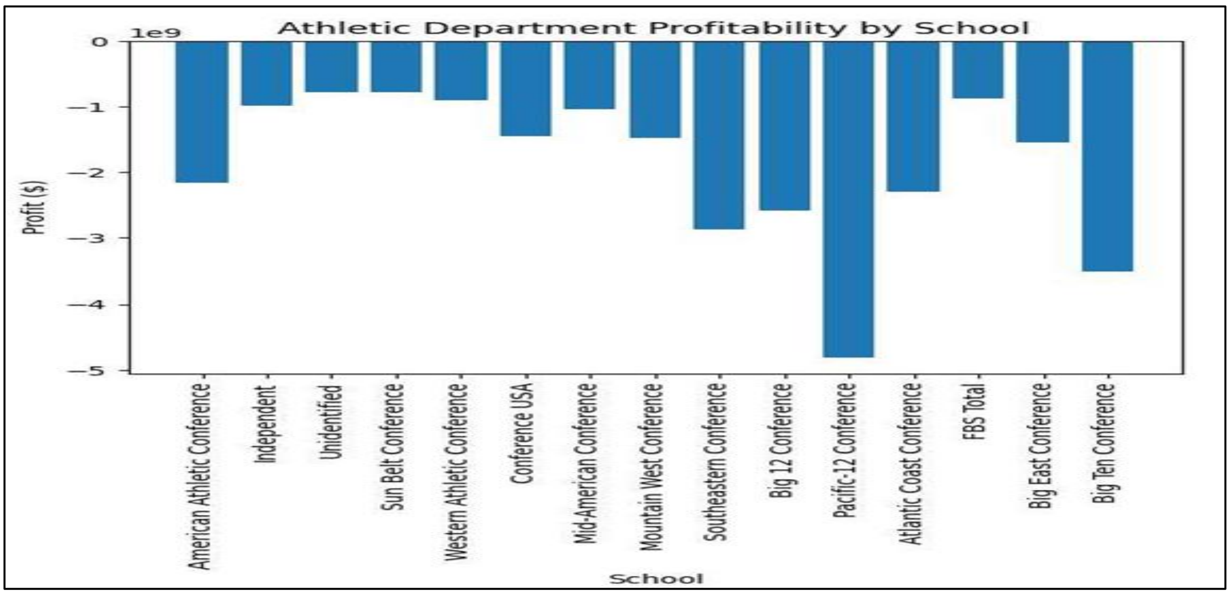


Figure 5: Athletic department profitability by school

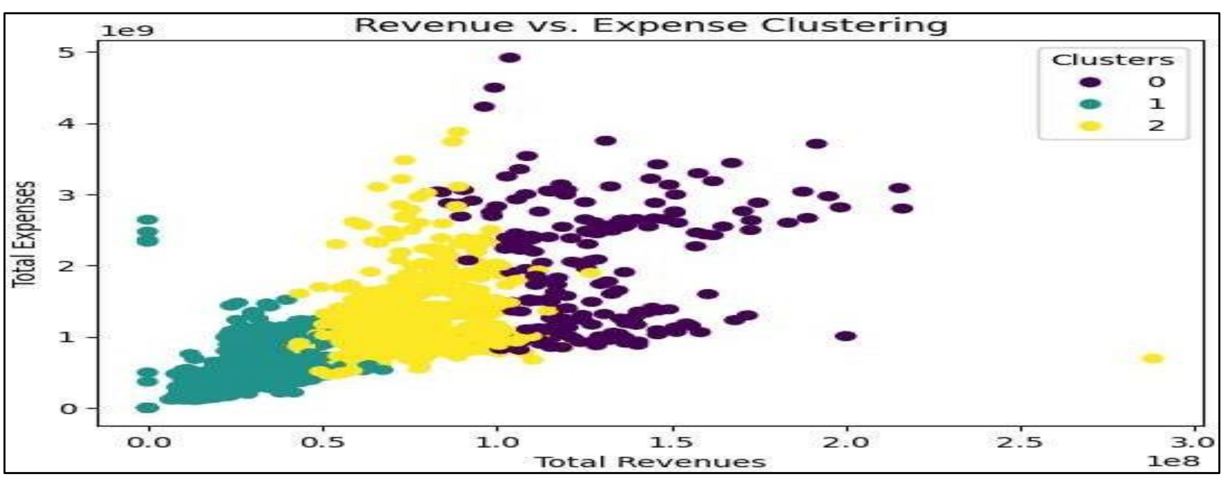


Figure 6: Revenue Vs expense

6th Experiment: To determine areas where our clubs' athletic department may need to concentrate its efforts to strengthen its competitive position within the conference, we compare revenue and spending categories against other clubs in the conference. Our club can use this data to determine whether we need to boost revenue in specific categories, cut costs in specific areas, or do both to improve our position in the conference. This research can also show how our club's athletic department is doing in comparison to other universities at the conference, which can help the department make strategic decisions and determine its top priorities.

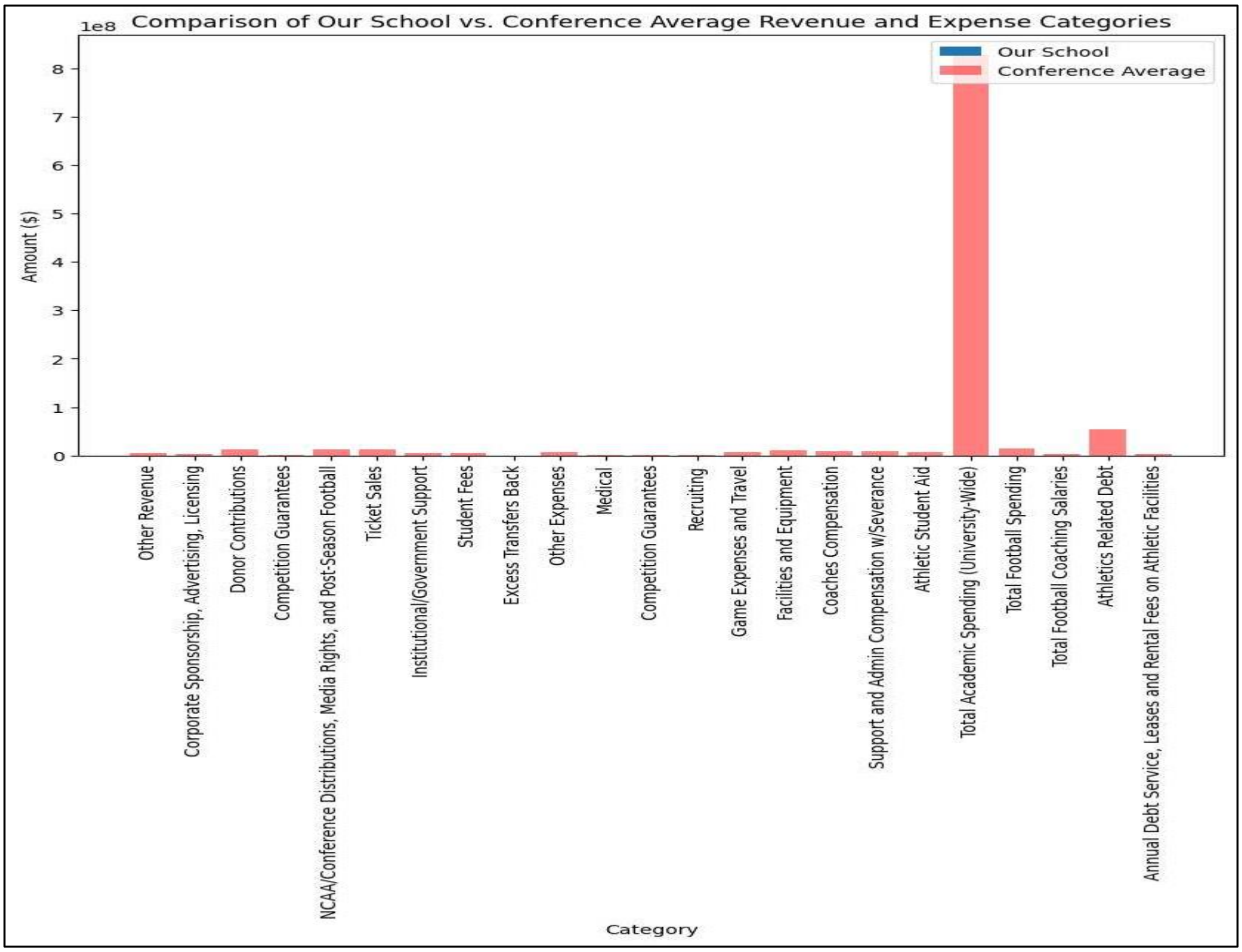


Figure 7: Comparison of schools

7th Experiment: A time series plot of total income over time is shown in the first graph, which gives a general idea of the trend and seasonality of the data. By breaking down the time series into its trend, seasonal, and residual components, the second graphic makes it easier to spot patterns and oscillations in the data.

The third plot displays the time series' autocorrelation and partial autocorrelation functions, which are helpful in choosing the right time series model parameters.

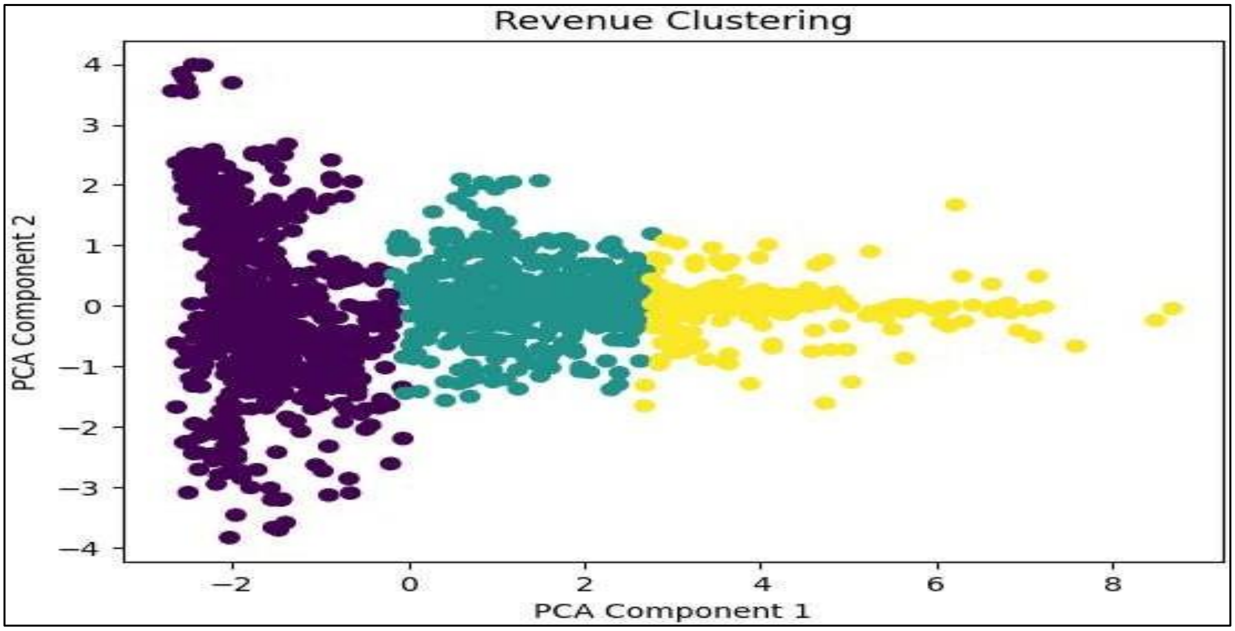


Figure 8: Revenue clustering

8th Experiment: The fourth portion predicts future revenue by fitting a seasonal ARIMA (SARIMA) model to the time series. A group of time series models known as SARIMA models also include seasonality and both moving average (MA) and autoregressive (AR) components. The values that are anticipated can be utilized to guide planning and decision-making for upcoming business operations.

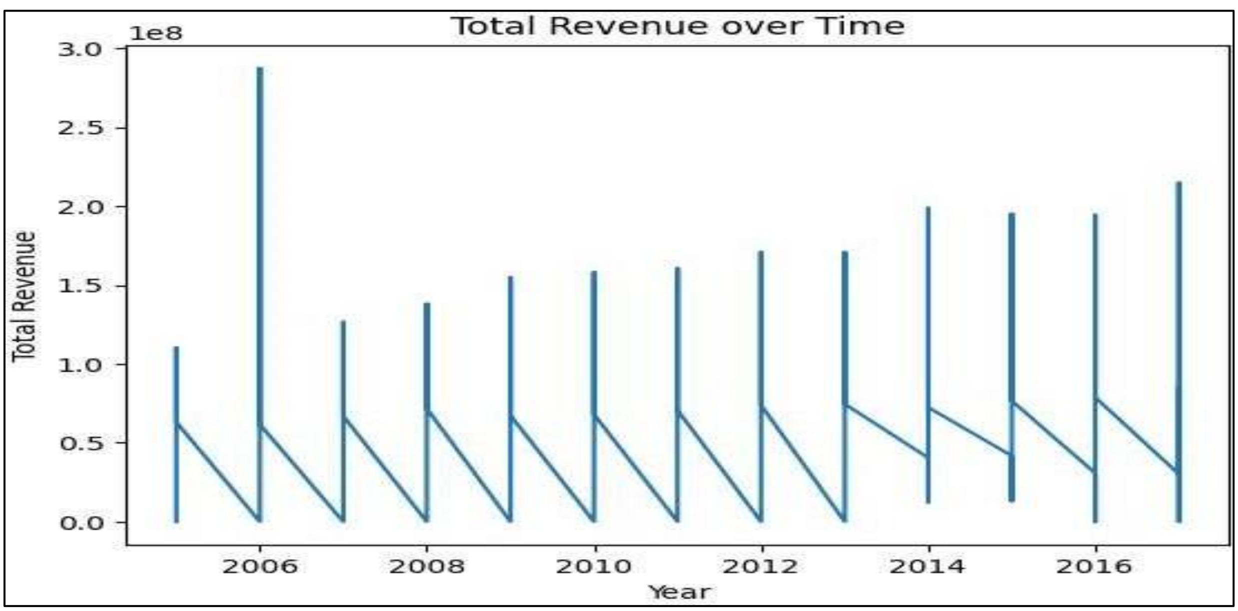


Figure 9: Total revenue over time

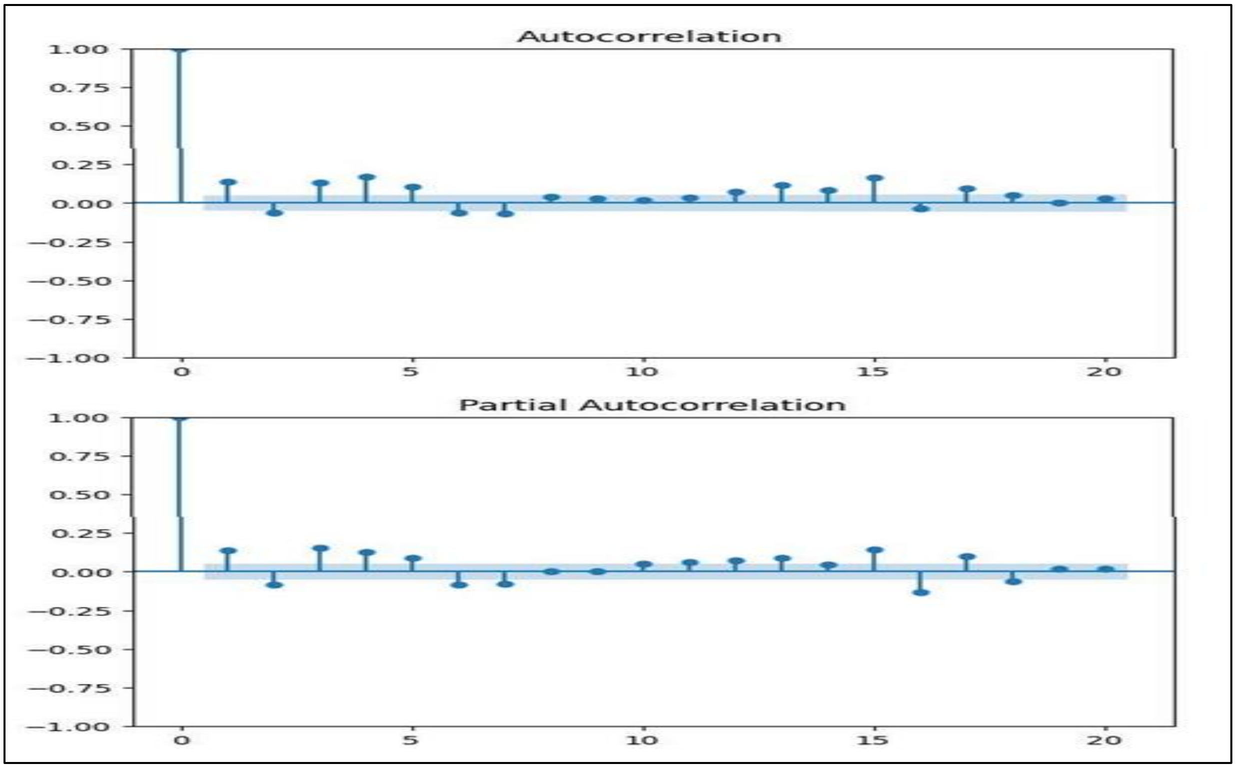


Figure 10: Autocorrelation

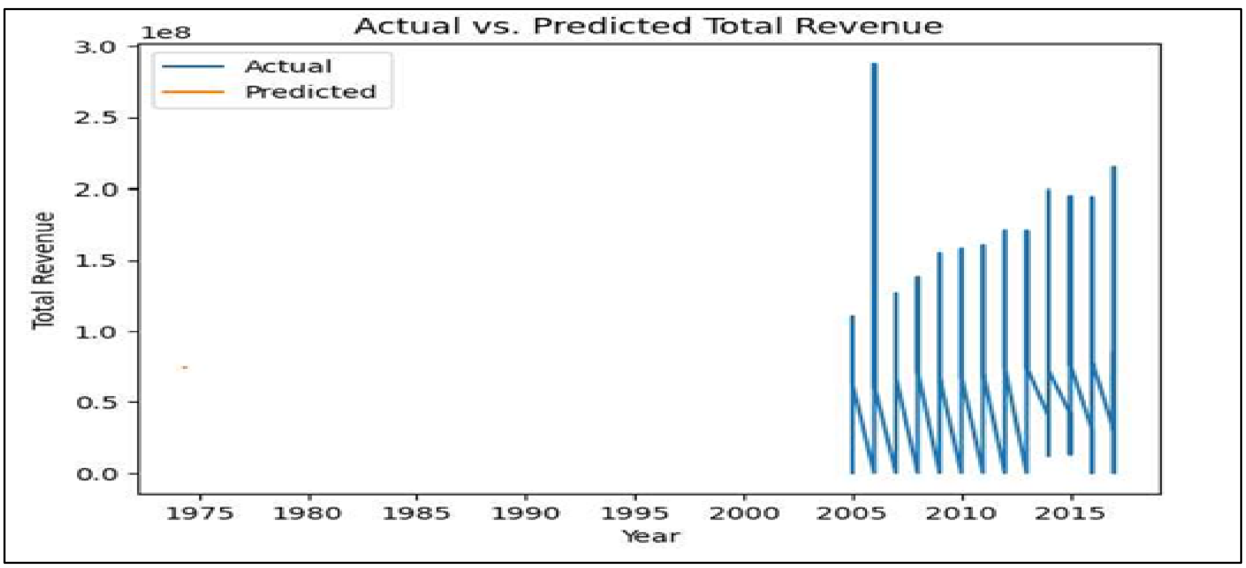


Figure 11: Actual vs predicted revenue.

4.2 Team Management & Player Performance analysis

9th experiment: Analysing Formation data to understand which formations yielded the result.
This can be divided in 2 categories. And one would be the formation data of the team of over 2+ years.

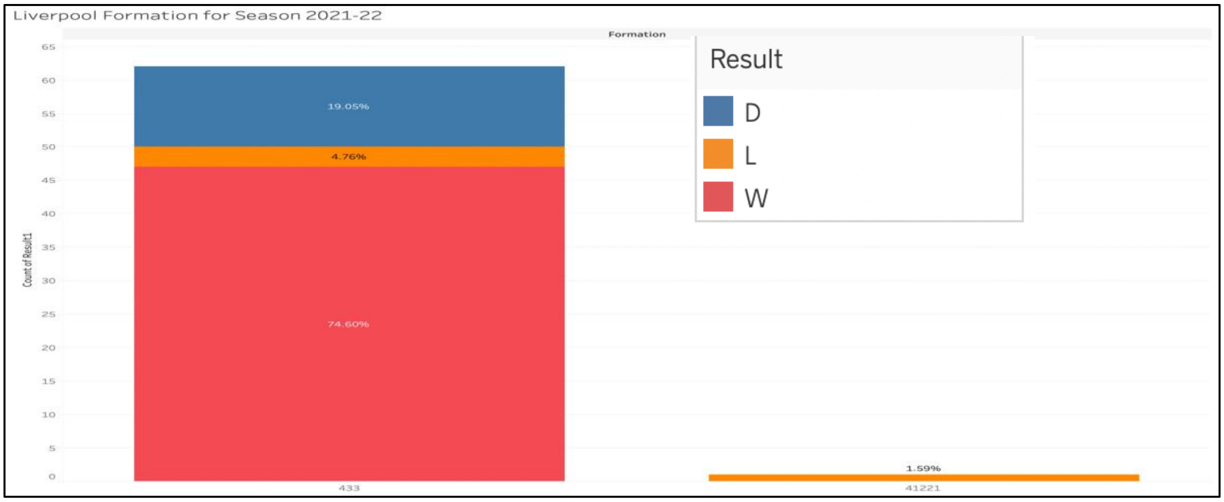


Figure 12: Formation for the season

Above distribution shows the usage of Team formation of Liverpool for the season 2021-22. It can observe that, there were only 2 distinct formations used of which standard 4-3-3 was subsequently highly effective. Whooping 75% of the matches played with that formation yielded in a result favour of our team. To assess the reason for success of 2021-22 season, we took a look at the squad and their metrics.

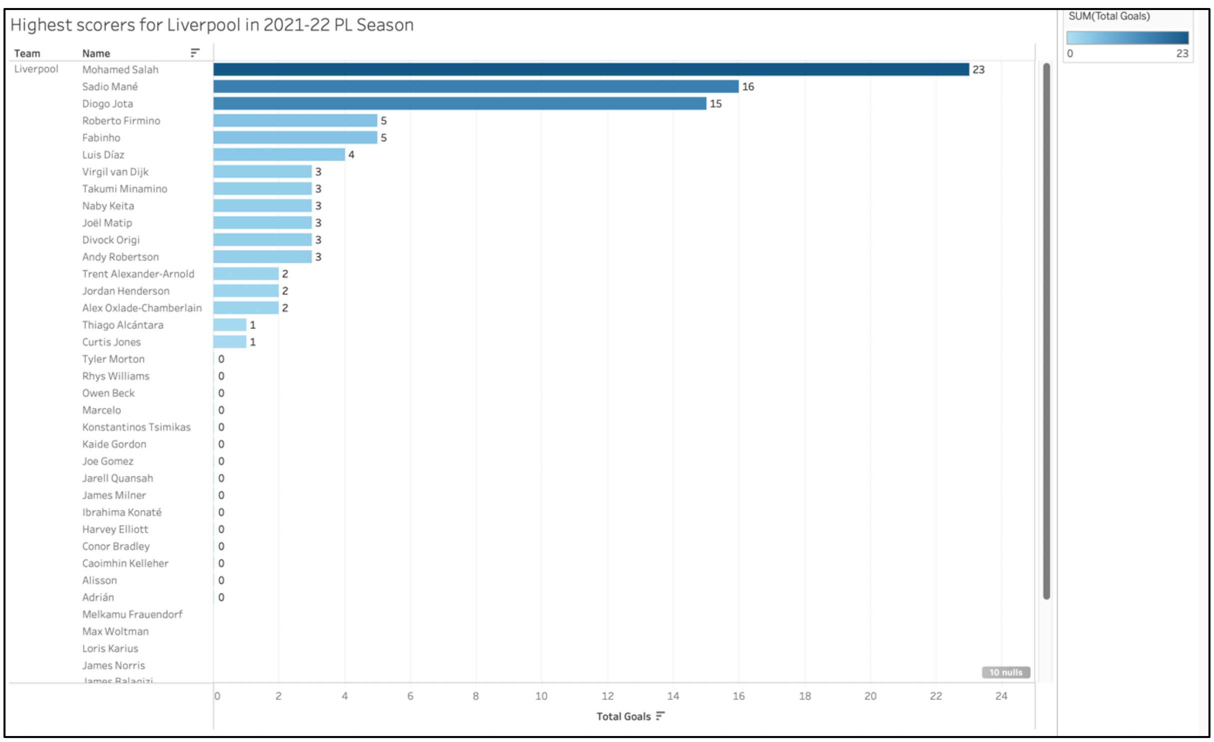


Figure 13: Highest scorers for Liverpool

We can see from the above goal scoring chart that; the forwards Mo. Salah and Sadio Mane were very crucial and provided the necessary width and pace to exploit the opponents defence line. However, this also proves a heavy reliance on specially trained forwards like Mane and Salah to prove the point the of difference during a bottle neck situation. As we observe the contribution from other team members, the midfield and defence have combined 34 nettings for the team. As against if we take a look at our rivals Man city goal scoring distribution it appears to be evenly distributed.

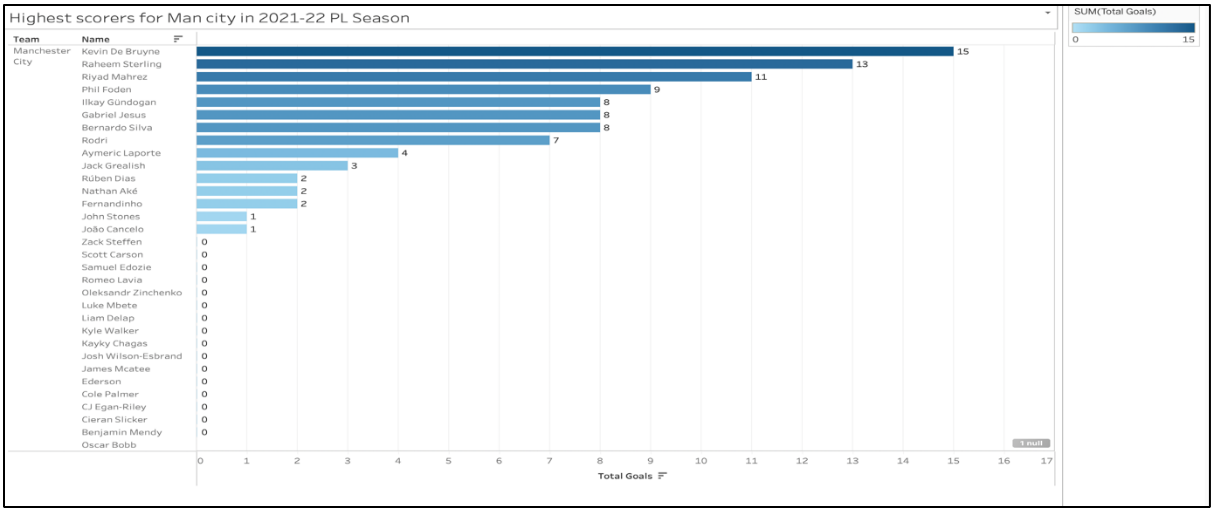


Figure 14: Highest scorers for Man city

This indicated that the set-pieces needs to be more converted as our play is mostly counter attacking. The midfield is understandably involved in conversion of defence to attack, the set-pieces are one of the areas that can be looked upon.

Also, in season 2022-23, Sadio Mane was sold to Bayern Munich a German giant which created a huge hole in the wing area. Due to which Liverpool tried adapting with various formations. As it can be seen in below representation, 5 different formations were tried in the season 2022-23.

Figure 15: Liverpool formation for the season

Even though the arrival of Darwin Nunez was thought to be a booster, the netting rates of the players have drastically lowered. This can be seen in the following presentation.

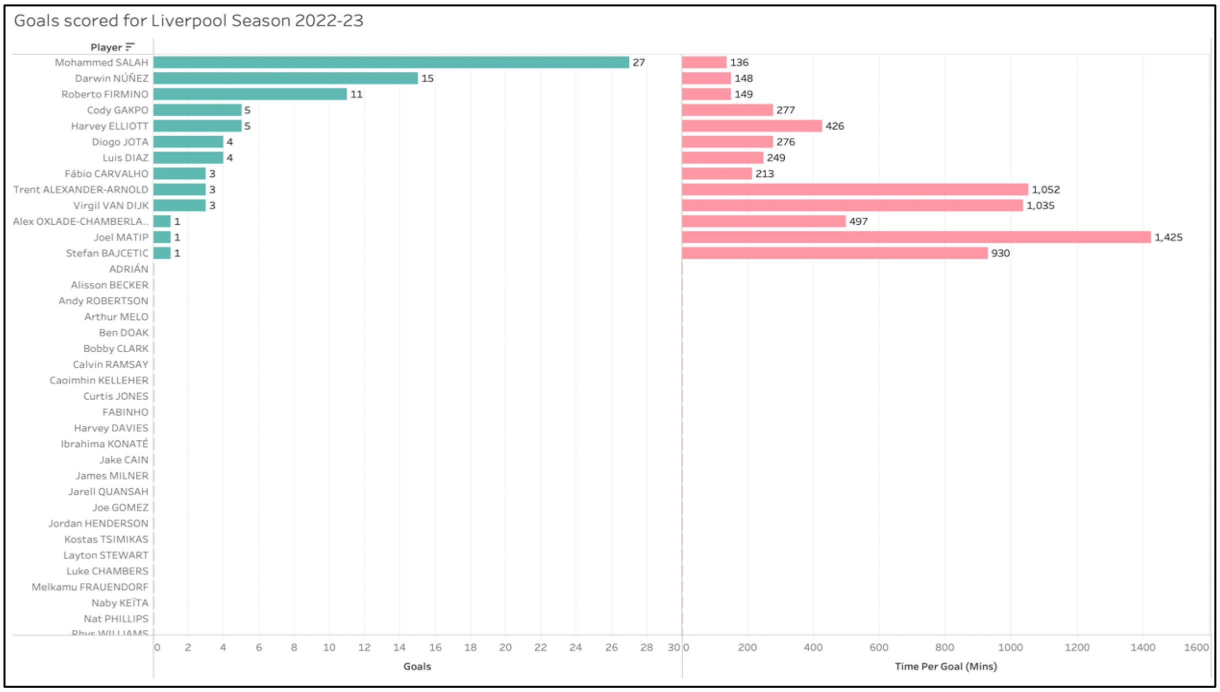


Figure 16: Goal scored for Liverpool.

10th experiment: Analysing Player stats for minutes & goals (Visualisation of the data)
Here we will be doing Player/Team data analysis and the Tableau presentation of the data to better understand the pov.

As shown in the above graphical representation, average time taken for each goal to be scored by the Liverpool players is more than the average time by rest of the top teams and especially rivals Man city.

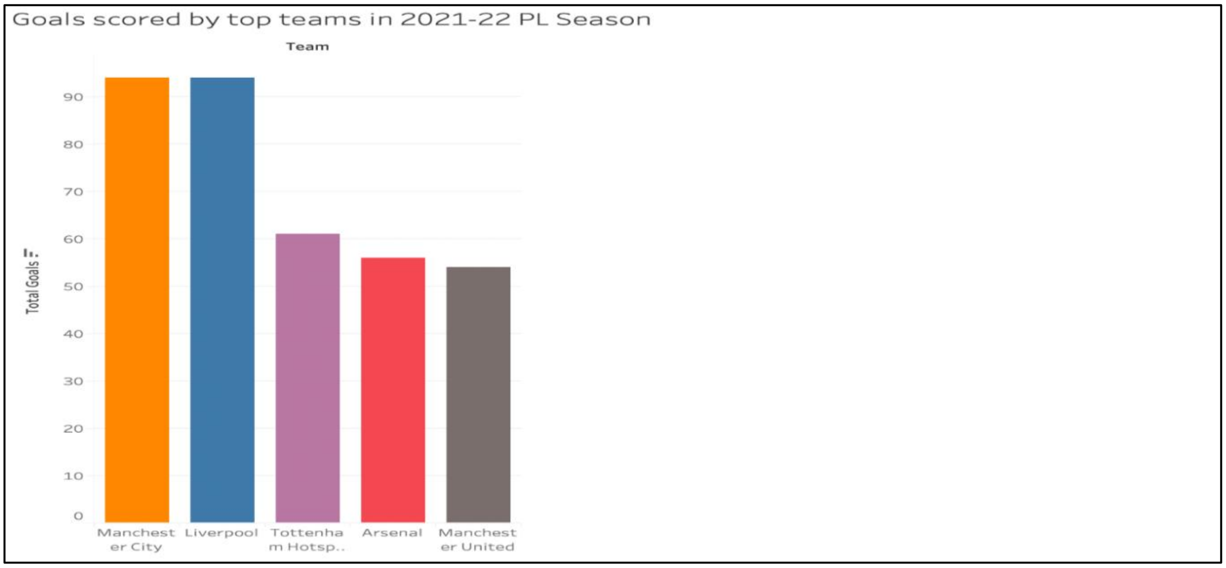


Figure 17: Goals scored by top teams.

As we can see that 2021-22 was a good season for Liverpool and players really pushed for goals to stay competitive on goal scoring record.

11th experiment: SWOT analysis of the team. This is a presentation of the team data based on the team data gathered. Here we are trying to see that what really changed in 2022-23 season for Liverpool. After the departure of Sadio Mane, the winger slot did not provide the breakthroughs that Liverpool was expecting and defending for opponents eased a bit. Even after trying out number of formations, as shown in earlier representations, team still produced mixed results. Adding to it, the defence line could not maintain the par that it had set in previous seasons. This can be seen from the FIFA ratings for both seasons compared to rivals Man city.

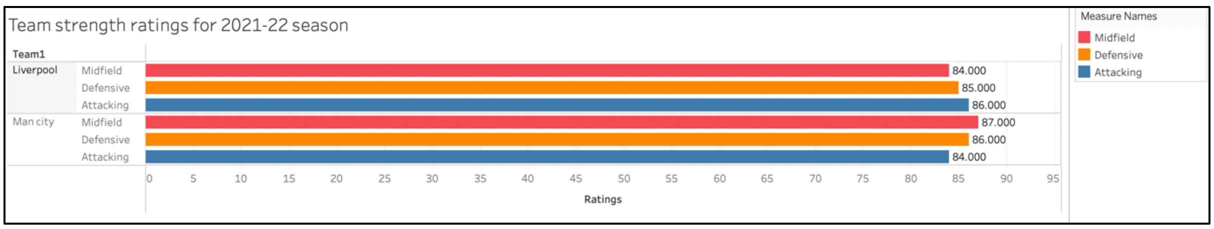


Figure 18: Team strength rating for 2021-22

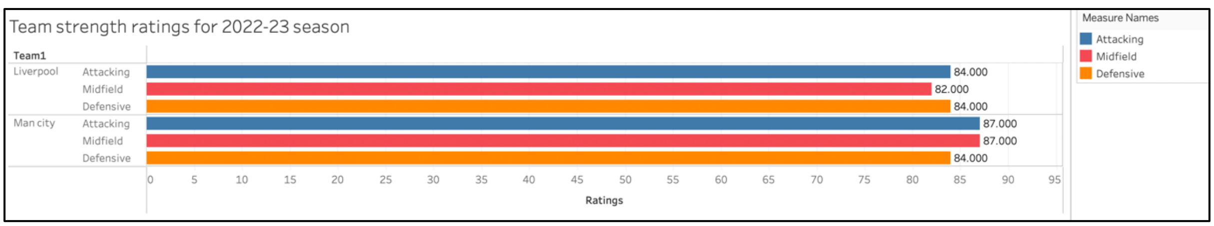


Figure 19: Team strength rating for 2022-23

It can be seen that defensive ratings have been dropped by a point and overall ratings have also been affected. If we take a look at goals conceded by teams' side by side, we can understand the disparity between this defensive gap.

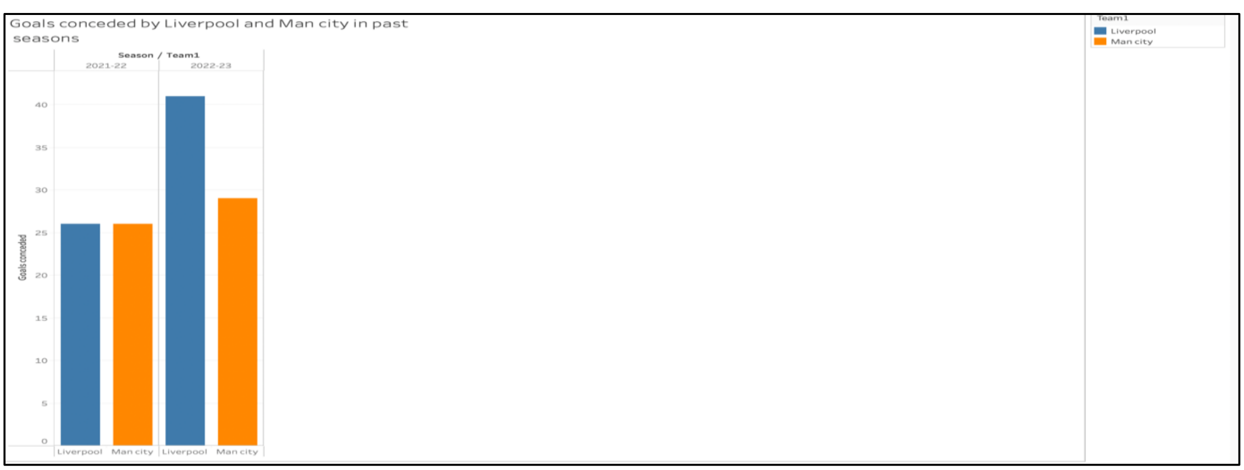


Figure 20: Goals conceded by Liverpool and Man city in past.

Liverpool has a substantial number of goals that were conceded in the season 2022-23. 43 times the opponents have breached the Liverpool backline as compared to 26 times in its previous season. Whereas Man City have managed to maintain their defence record with just 29 goals conceded in the current season.

12th experiment: Popularity and revenue impact. We gazed through the revenue to check whether poor form of their team would affect the Liverpool fans. But revenue seems to be up, and this shows belief in the team which is crucial for players to be backed up.

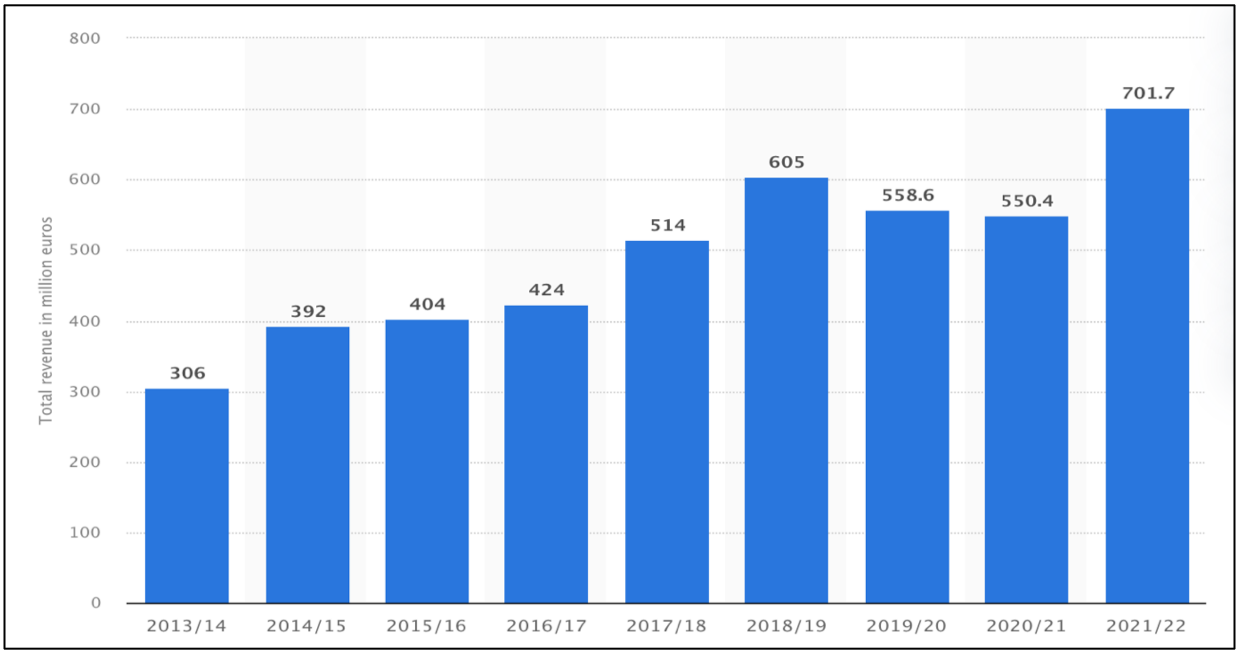


Figure 21: Revenue

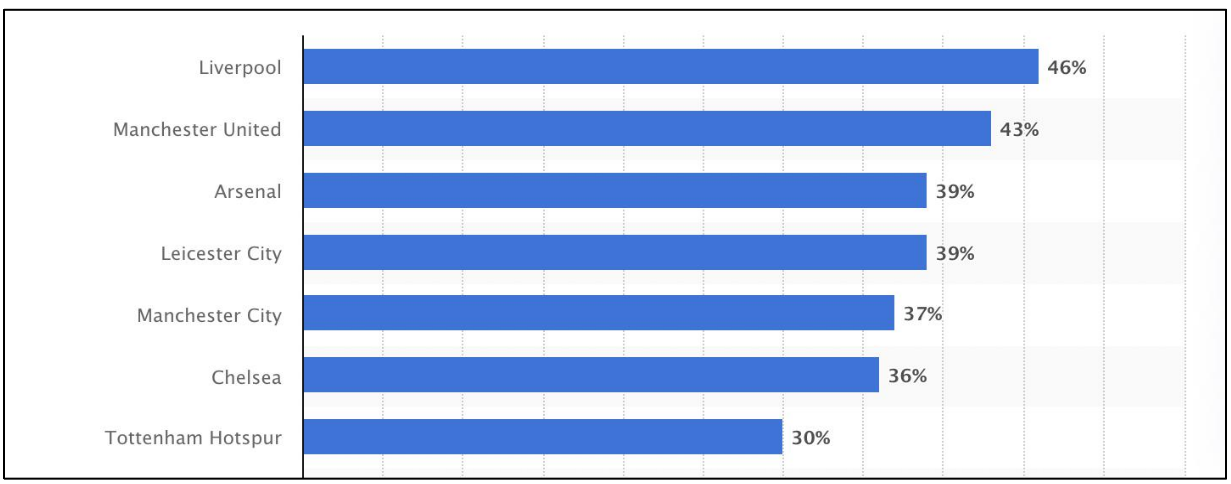


Figure 22: Fan following.

Liverpool still enjoys majority of the fan following across social media platforms and maximum support for the club in the UK. Around half of the UK football fan followers, support Liverpool club and this reflects better on the revenue sheet for merchandise.

4.3 Club Management and Administration analytics

13st Experiment: Descriptive analysis- distribution of the sponsorship budgets.
The first plot creates a histogram of the "Budget" variable, showing the distribution of the sponsorship budgets.

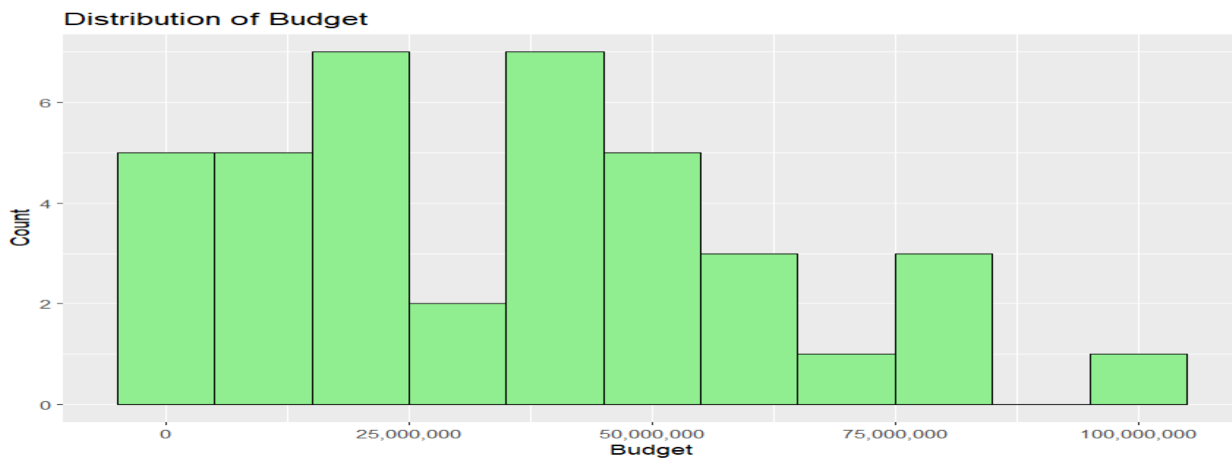


Figure 23: Distribution of budget

Using the histogram gives a clear and concise understanding of the budget. We can easily see the maximum and minimum range of budget. We can visualize that 6 sponsors have budgets of around 25000000 and other 6 have around 50000000, which gives a clear understanding and might require further investigation to explore more suitable options for revenue. Area band between 75000000-100000000 have minimum (0-3) sponsors. Overall histogram represents a diverse range of budget allocation with number of sponsors increases if budget < 50000000.

14th Experiment: Budget vs. Annual Budget

Scatter plot of "Budget" vs. "Annual Budget", coloring the points by the "Sponsorship level". This plot shows how the budgets and annual budgets are related by sponsorship level.

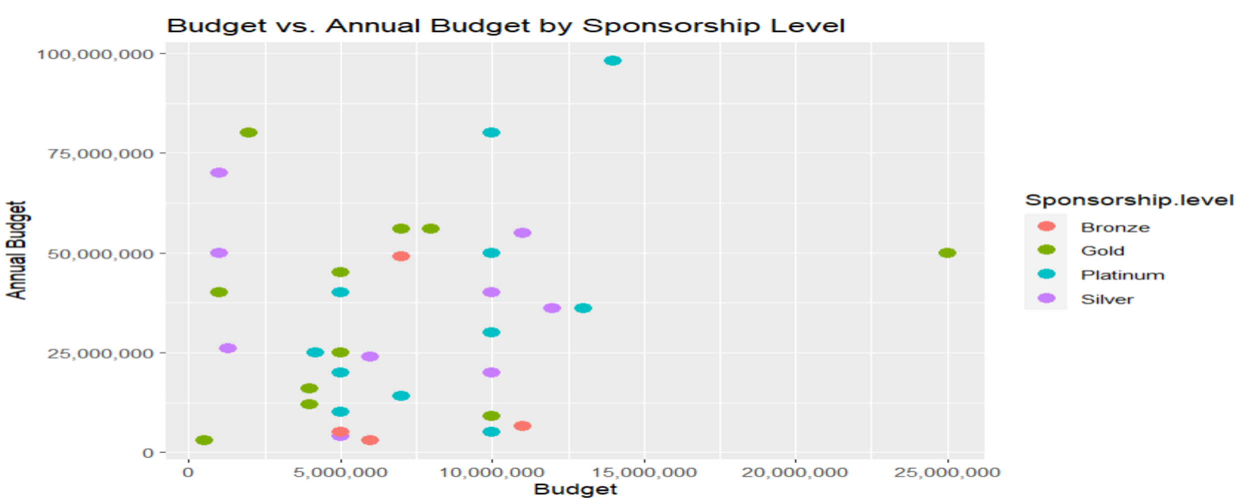


Figure 24: Budget vs. Annual Budget

With the scatter plot we can visualize that budget is directly proportional to annual budget. As the annual budget increases the budget also tends to increase. With a gold sponsorship level with a budget of 25000000, a particular sponsor should focus on that specific outlier to check out or require particular investigation in that area. As we can see there is a dense area within a range of annual budget of 50000000 and budget of 10000000,it can be helpful for decision maker that this is the specific lucrative market a company should target.

15th Experiment: Correlation Analysis

The dataset's correlation matrix displays the correlation coefficients between several variables. The intensity and direction of the linear link between two variables are measured by correlation coefficients. A coefficient of 1 denotes a positive correlation, a coefficient of 0 denotes no connection, and a coefficient of -1 denotes a flawless negative correlation.

The correlation matrix shows a positive connection (r = 0.39) between the sponsorship term and success rate and a moderately positive correlation (r = 0.32) between the sponsorship term and budget. There is a positive correlation (r = 0.15) between the budget and market share. However, there is a weak correlation between the yearly budget and other factors.

A bigger budget and higher success rate are related to a larger market share. Insight into the correlation between factors is limited by the poor association between the sponsorship year and other variables.

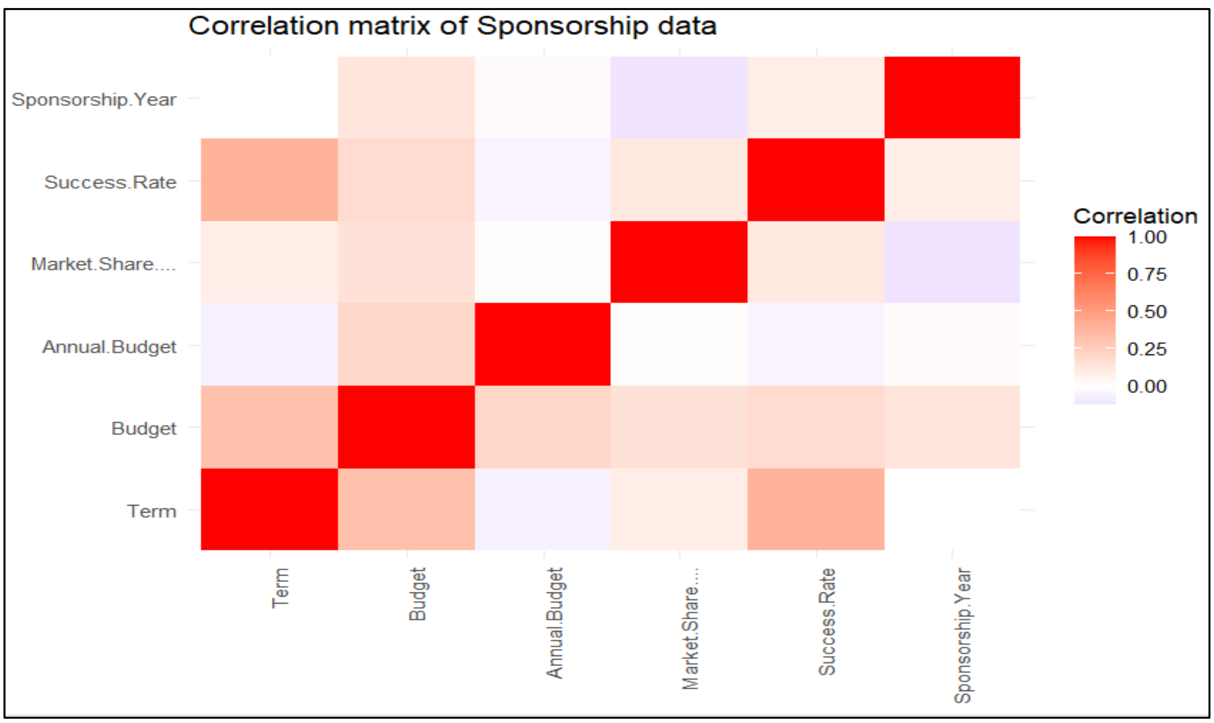


Figure 25: Correlation Analysis

4.4 Predictive analytics in sports for team and detailed players analysis

16th Experiment: Do half-time result match the end in general.
We will try to see that in how many games did the half-time result match the end in general? How has this changed over the years?

```
[15] liverpool_data=Analysis_file.loc[Analysis_file['HomeTeam']=='Liverpool']

liverpool_data.tail()
```

	Season	DateTime	HomeTeam	AwayTeam	FTHG	FTAG	FTR	HTHG	HTAG	HTR	Referee	HS	AS	HST	AST	HC	AC	HF	AF	HY	AY	HR	AR
11026	2021-22	2022-02-10T19:45:00Z	Liverpool	Leicester	2	0	H	1.0	0.0	H	C Kavanagh	22.0	5.0	11.0	1.0	7.0	6.0	11.0	7.0	1.0	0.0	0.0	0.0
11043	2021-22	2022-02-19T15:00:00Z	Liverpool	Norwich	3	1	H	0.0	0.0	D	M Dean	29.0	6.0	8.0	1.0	9.0	4.0	7.0	8.0	0.0	1.0	0.0	0.0
11050	2021-22	2022-02-23T19:45:00Z	Liverpool	Leeds	6	0	H	3.0	0.0	H	M Oliver	23.0	3.0	15.0	2.0	4.0	0.0	5.0	11.0	0.0	3.0	0.0	0.0
11067	2021-22	2022-03-05T17:30:00Z	Liverpool	West Ham	1	0	H	1.0	0.0	H	J Moss	22.0	13.0	5.0	5.0	4.0	6.0	7.0	7.0	2.0	1.0	0.0	0.0
11092	2021-22	2022-04-02T12:30:00Z	Liverpool	Watford	2	0	H	1.0	0.0	H	S Attwell	20.0	5.0	3.0	2.0	9.0	3.0	9.0	8.0	1.0	1.0	0.0	0.0

```
# how many rows have the values match over total rows: 60.4%
# half-time results predict full-time results a majority of the time
Home_win_percent_at_half_time=liverpool_data.loc[liverpool_data['FTR'] == liverpool_data['HTR']].shape[0] / liverpool_data.shape[0] * 100

[18] Home_win_percent_at_half_time

60.07259528130672
```

The result shows that the 60 percent of home games ended same as that of half-time result. Over the season's following trend was observed for home results.

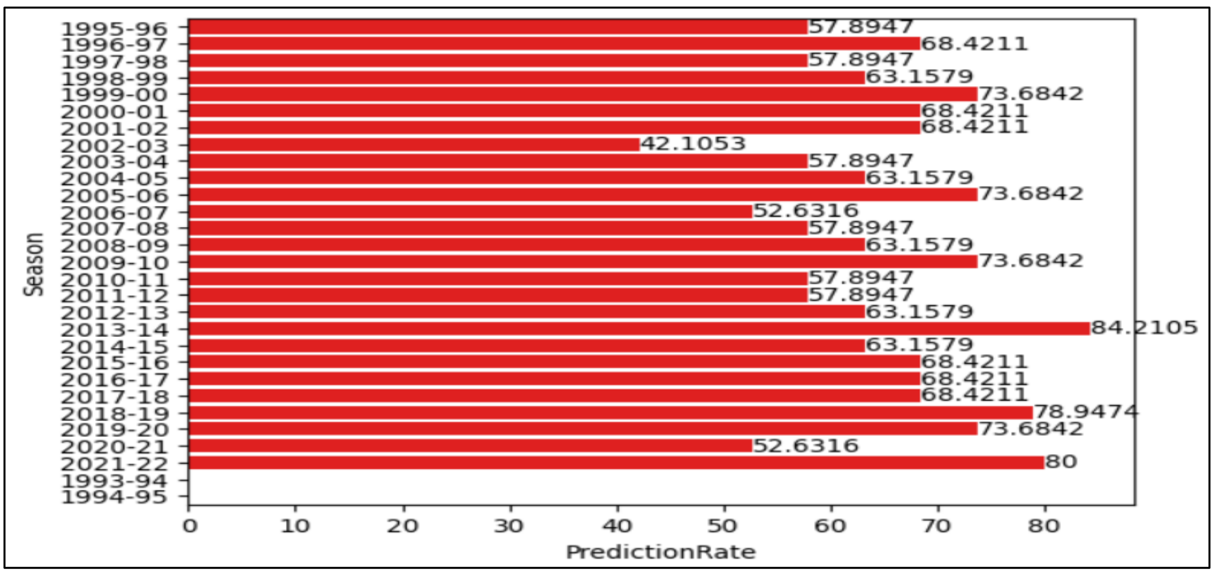


Figure 26: Prediction

For away results 57.78 times the half time results matched the full-time score.

```
liverpool_data_Away=Analysis_file.loc[Analysis_file['AwayTeam']=='Liverpool']
Away_wins_halftime=liverpool_data_Away.loc[liverpool_data_Away['FTR'] == liverpool_data_Away['HTR']].shape[0] / liverpool_data_Away.shape[0] * 100

[21] Away_wins_halftime

57.789855072463766
```

Figure 27: Half time prediction

The following trends were followed over the seasons for away form at half-time wins.

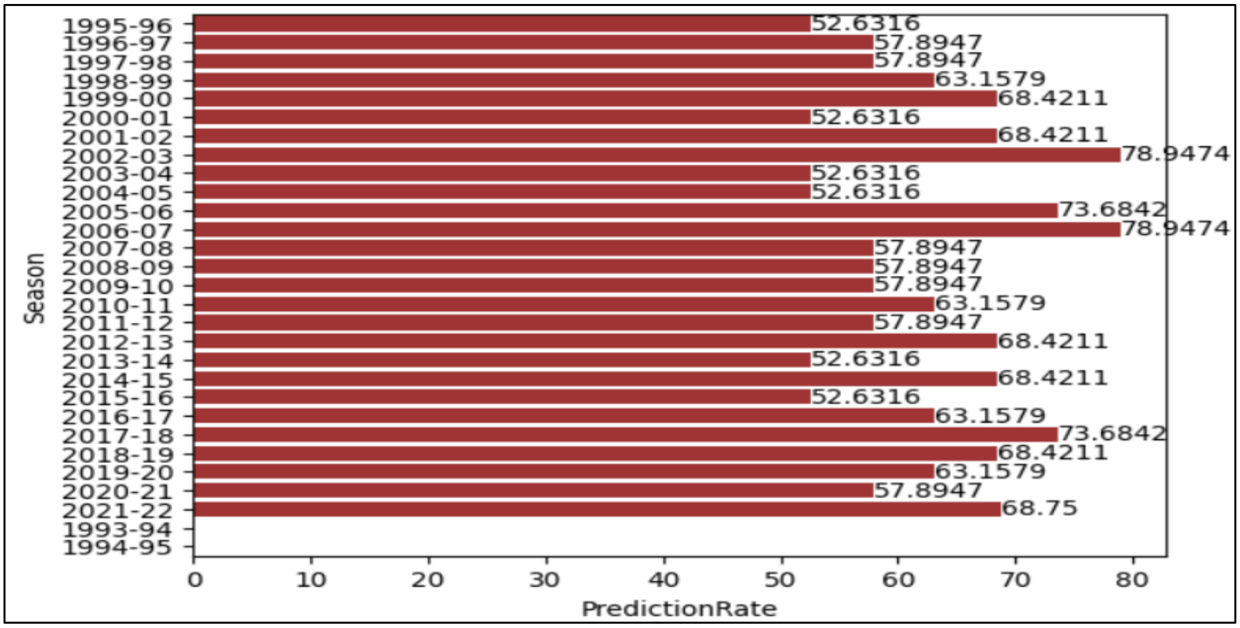


Figure 28: Prediction rate

17th Experiment: Home and away win percentages for Liverpool.

```
home_win=100*round((liverpool_data.loc[liverpool_data['FTR']=='H'].groupby('HomeTeam')['FTR'].count()/liverpool_data.groupby('HomeTeam')['FTR'].count()),3)
home win
```

Home win percentage = 63

```
Away_win=100*round((liverpool_data_Away.loc[liverpool_data_Away['FTR']=='A'].groupby('AwayTeam')['FTR'].count()/liverpool_data_Away.groupby('AwayTeam')['FTR'].count()),3)
Away win
```

Away win percentage = 43.5

18th Experiment: Constructing points table from the given data and getting information about goals scored, goals conceded and total points at the end of each season.

To retrieve the above information, we must do some calculations on the given data for home and away results.

The home and away results stats table is as follows.

	Season	HomeTeam	HP	HW	HL	HD	HG Against	HG For	HG For(Half Time)	HG Against(Half Time)
2	1995-96	Liverpool	19	14	1	4	13	46	5	20
3	1996-97	Liverpool	19	10	3	6	19	38	5	20
4	1997-98	Liverpool	19	13	4	2	16	42	5	14
5	1998-99	Liverpool	19	10	4	5	24	44	12	20

Figure 29: Home stats

	Season	AwayTeam	AP	AW	AL	AD	AG For	AG Against	AG For(Half Time)	AG Against(Half Time)
2	1995-96	Liverpool	19	6	6	7	24	21	12	12
3	1996-97	Liverpool	19	9	5	5	24	18	9	10
4	1997-98	Liverpool	19	5	5	9	26	26	13	9
5	1998-99	Liverpool	19	5	10	4	24	25	14	10
6	1999-00	Liverpool	19	8	5	6	23	17	12	11

Figure 30: Away stats

Later we concatenated these into Data frames to create a table for all seasons for Liverpool using the following python code.

```
pt_table.head()
```

	Season	Team	HP	HW	HL	HD	HG Against	HG For	HG For(Half Time)	HG Against(Half Time)	AP	...	AG Against	AG For(Half Time)	AG Against(Half Time)	Total Played	Total Wins	Total Loss	Total Draw	GF	GA	GD	Total_Points
0	1995-96	Liverpool	19	14	1	4	13	46	5	20	19	...	21	12	12	38	20	7	11	70	34	36	71
1	1996-97	Liverpool	19	10	3	6	19	38	5	20	19	...	18	9	10	38	19	8	11	62	37	25	68
2	1997-98	Liverpool	19	13	4	2	16	42	5	14	19	...	26	13	9	38	18	9	11	68	42	26	65
3	1998-99	Liverpool	19	10	4	5	24	44	12	20	19	...	25	14	10	38	15	14	9	68	49	19	54
4	1999-00	Liverpool	19	11	4	4	13	28	8	11	19	...	17	12	11	38	19	9	10	51	30	21	67

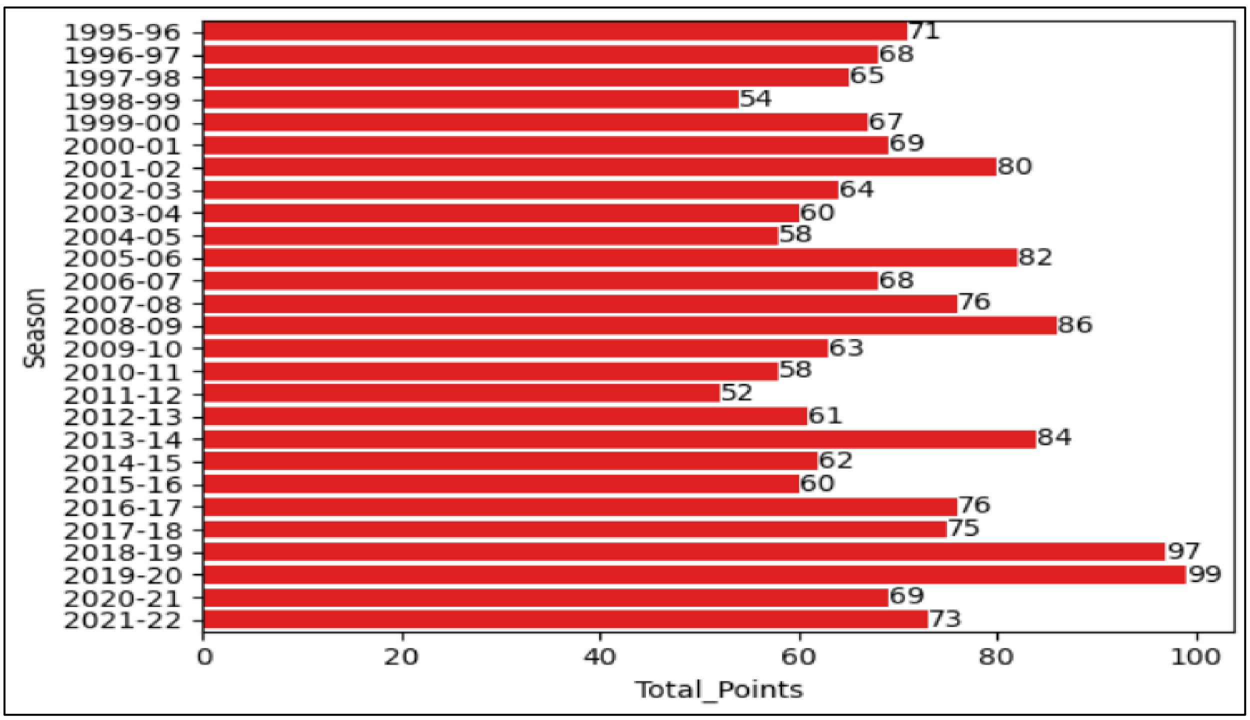


Figure 31: Points table over the years

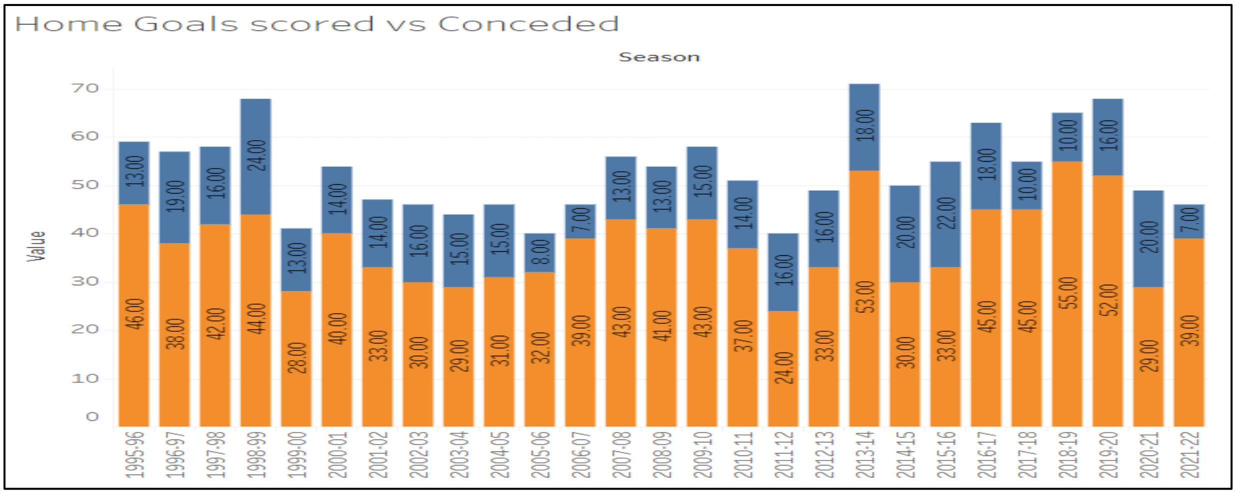


Figure 32: Home goals scored vs conceded.

19th Experiment: Can corners predict goals or results.

Generally, teams win corners from attacking moves. Teams tend to win corners when they are on their front foot. So, let's see if there are patterns, we can find with them. Do teams with more corners tend to win? Do they tend to score more goals? Do they tend to have more shots? Intuitively, one would think there's some relationship here.

Check results first; let's give a 3-corner leeway. Being within 3 corners means either a draw-or-win for the team with more is still 'accurate'.

Even with that bit of leeway, using corners to determine the winner is accurate only 35.7% of the time.

How about goals, though? Let's plot goals against corners and see what we find.

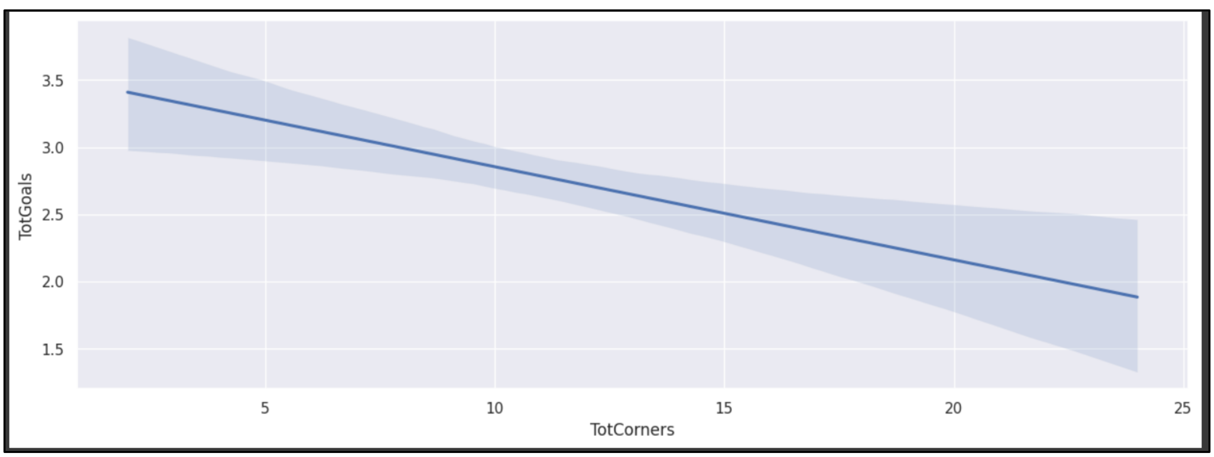


Figure 33: corners predict goals or results.

In general, we see that games with many corners aren't often high-scoring games.

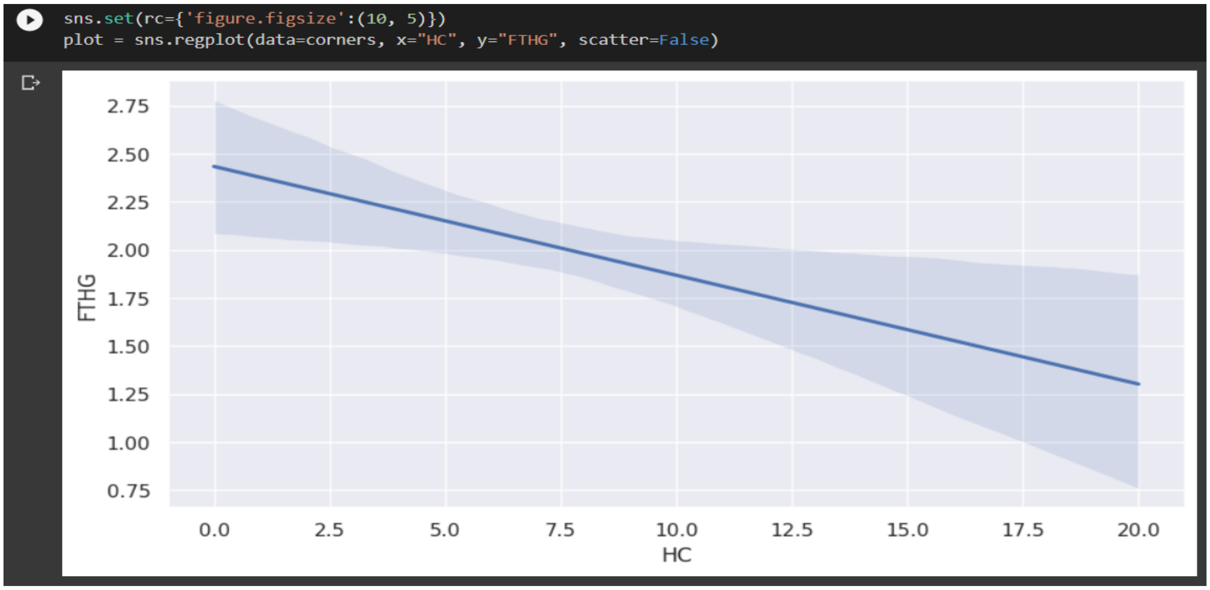


Figure 34: For Home Goals and Home corners the team is not able to convert them into goals.

20th Experiment: Neural Network Approach to predict full time result.

First, we created a copy of the data to work on. After we dropped the referee column and converted the datetime to weekday as the matches played on a particular day can affect the match results. As the inputs should be in binary format, categorical variables were converted to Binary. Home team, away team, half time results and match day were taken as input parameters. And Full Time Result was the y test variable which we will be predicting. TensorFlow Keras model was used for prediction. Activation was set to relu and then the model was fitted.

After fitting the model, we observed that it was overfitting after certain epochs as shown in the graph below.

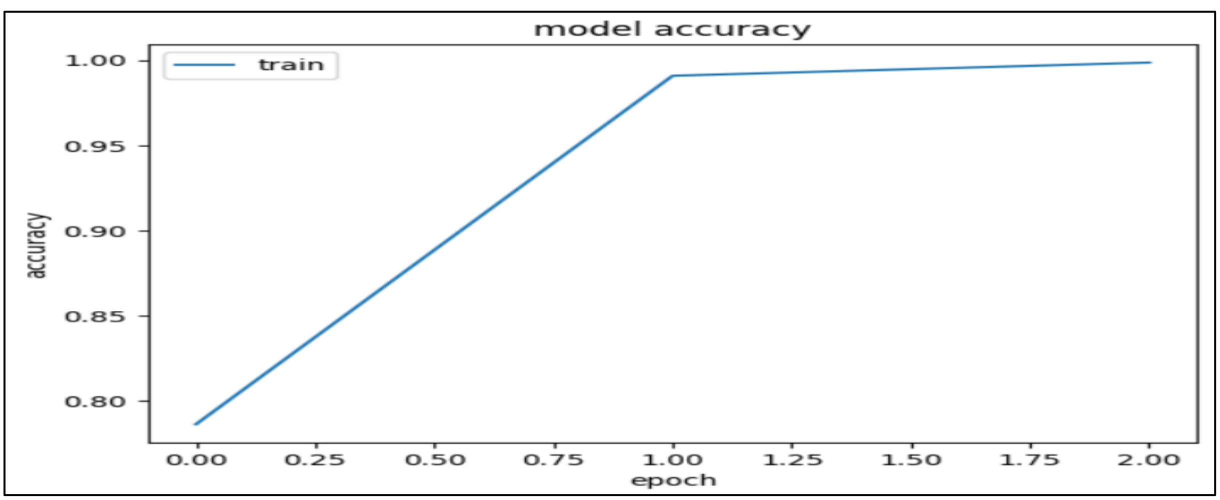


Figure 35: Neural Network Approach to predict full time result.

Hence, we moved to the next model that is Random Forest Classifier.

21st Experiment: Random Forest Classifier Approach to predict full time result.

The (random forest) algorithm establishes the outcome based on the predictions of the decision trees. It predicts by taking the average or mean of the output from various trees. Increasing the number of trees increases the precision of the outcome.

We imported it from sklearn.

The input parameters were set to home team, away team, Year, match Month and half-time score.

The prediction variable was full time result.

H = Home team win, A= Away team win, D= Draw

After splitting the data into training and tests sets, we fitted the model, and the depth was set to be max.

Following were the results of the classification report.

```
from sklearn.metrics import classification_report, accuracy_score

val_predictions = pl_results_model.predict(val_X)

print(accuracy_score(val_y, val_predictions))
print(classification_report(val_y, val_predictions))
```

```
0.6134222919937206
              precision    recall  f1-score   support

           A       0.62      0.61      0.62       716
           D       0.43      0.30      0.35       671
           H       0.67      0.79      0.73      1161

    accuracy                           0.61      2548
   macro avg       0.57      0.57      0.57      2548
weighted avg       0.59      0.61      0.60      2548
```

The model was 61.34 % accurate.

The following parameters were given as inputs.

Home team—Chelsea

Away team – Liverpool

Year—2023

Month – May

Half time score home team – 1

Half time score away team – 1

```
[35] print(predict_result(pl_results_model, 'Chelsea', 'Bolton', 2022, 5, 1, 1))

     ['H']
```

The result is H, which means Chelsea will win that game.

Further we also predicted the probabilities of the results.

```
print(predict_probabilities(pretrained_models, 'Liverpool', 'Man United', 2024, 4, 3, 4))
{'H': 0.8700000000000007, 'D': 0, 'A': 0.13000000000000003}
```

Here the Liverpool vs Man United match if its 3-4 at half time. The Probabilities are as above.

22nd Experiment: Detailed player's analysis.

Table 1: Player's profile

S. No.	Particular	Results
1	Average Age	25.13
2	Average Height	71.99
3	Average weight	166.9
4	Average min on field per match	54.14
5	Average goal	1.7

23rd Experiment: Player Radial Diagram of 4 top players

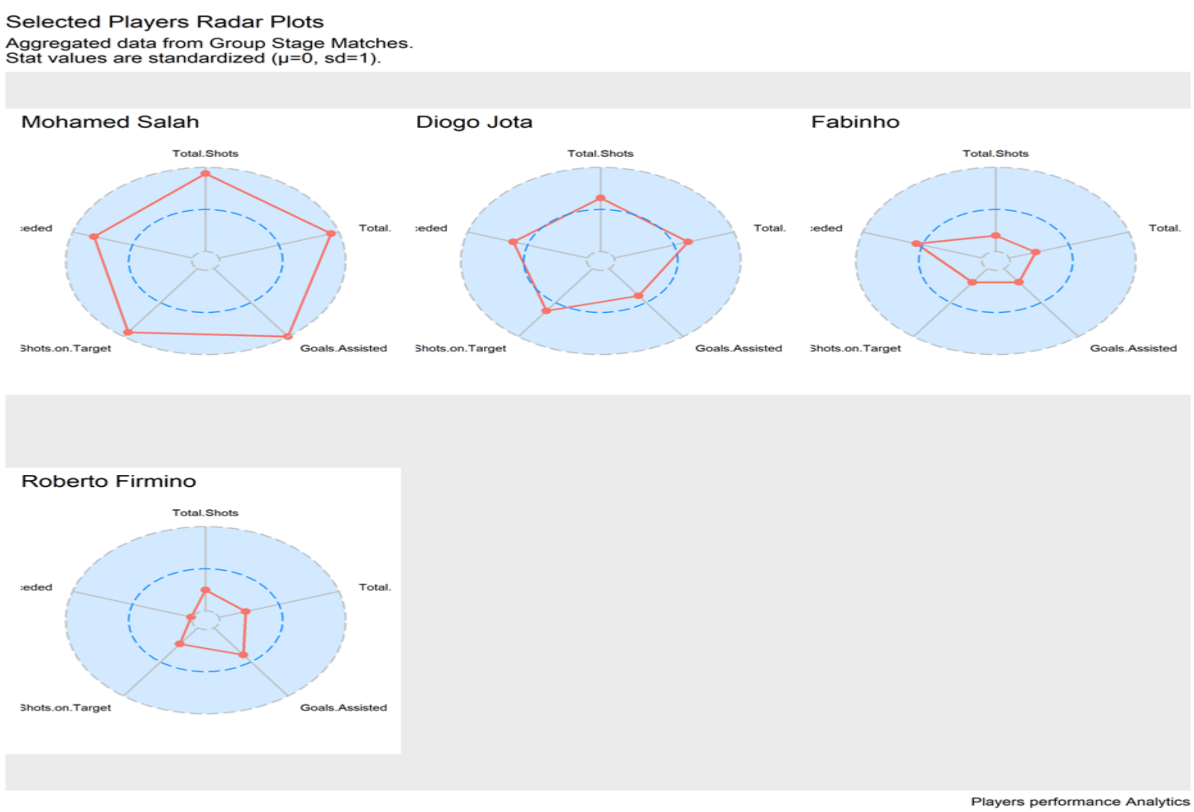


Figure 36: Player Radial Diagram of 4 top players

24th Experiment: Total shots vs total goal scored.

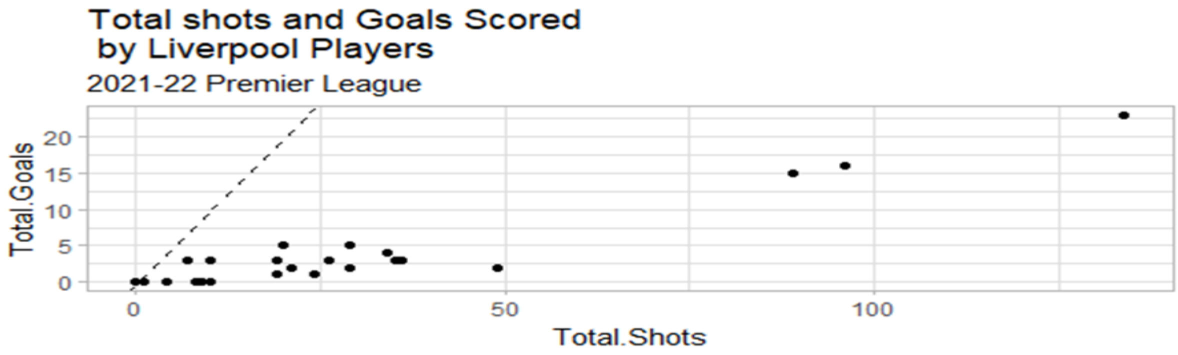


Figure 37: Total shots vs total goal scored.

25th Experiment: Do the number of goals depend on the physical strength of the player. Like height weight or age affect the goal scoring ability.

```
> # Fit the regression model
> model <- lm(all_stat$Total.Goals ~ Weight + Height + Age, data = all_stat)
> # Print the summary of the model
> summary(model)

Call:
lm(formula = all_stat$Total.Goals ~ Weight + Height + Age, data = all_stat)

Residuals:
    Min      1Q  Median      3Q     Max
-4.0262 -1.9073 -1.0527  0.4964 20.9406

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 20.093491   4.710810   4.265 2.42e-05 ***
Weight      -0.009855   0.014017  -0.703  0.48235
Height      -0.229250   0.085766  -2.673  0.00778 **
Age          0.005759   0.037320   0.154  0.87743
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.348 on 463 degrees of freedom
  (203 observations deleted due to missingness)
Multiple R-squared:  0.05112,	Adjusted R-squared:  0.04498
F-statistic: 8.315 on 3 and 463 DF,  p-value: 2.139e-05
```

This is a linear regression model that attempts to predict the Total Goals scored by soccer players based on their Weight, Height, and Age. The regression model equation is:

Total Goals = 20.09 - 0.0099 x Weight - 0.229 x Height + 0.0058 x Age

The coefficients indicate the effect of each variable on the dependent variable (Total Goals) while holding all other variables constant.

The intercept of 20.09 means that, on average, players with a weight of 0, height of 0, and age of 0 are expected to score 20.09 goals.

The coefficient for weight is -0.0099, indicating that for every one-unit increase in weight, the expected number of goals scored decreases by 0.0099 while holding height and age constant. However, the p-value for weight is 0.48, which is not statistically significant, suggesting that weight does not have a significant impact on the number of goals scored after controlling for height and age.

The coefficient for height is -0.229, which means that for every one-unit increase in height, the expected number of goals scored decreases by 0.229 while holding weight and age constant. The p-value for height is 0.0078, which is statistically significant, indicating that height has a significant effect on the number of goals scored.

The coefficient for age is 0.0058, indicating that for every one-unit increase in age, the expected number of goals scored increases by 0.0058 while holding weight and height constant. However, the p-value for age is 0.877, which is not statistically significant, suggesting that age does not have a significant impact on the number of goals scored after controlling for weight and height.

The multiple R-squared value of 0.0511 indicates that the three predictor variables (Weight, Height, and Age) explain only 5.11% of the variability in Total Goals. The adjusted R-squared value of 0.0449, which is slightly lower than the multiple R-squared, indicates that the model's fit is not improved by including more variables.

The F-statistic tests the overall significance of the model, and the p-value of 2.139e-05 indicates that the model is statistically significant, meaning that the predictor variables collectively have a significant effect on the dependent variable.

The residual standard error (RSE) is a measure of the average deviation of the actual Total Goals from the predicted Total Goals. The RSE value of 3.348 means that the model's predictions are typically off by around 3.348 goals.

5. Discussion

The act of deliberate contemplation is one that demands one's attention. It is a process of careful consideration and thoughtfulness that requires one to focus on a particular topic or issue. After conducting a thorough examination in the previous sections, a comprehensive understanding of the football club's financial situation, player achievements, and team administration has been obtained. These understandings can aid stakeholders in making informed choices regarding player transfers, scouting operations, promotional campaigns, and other related subjects.

Our financial research has yielded critical discoveries that can benefit the operational efficiency of football teams. In particular, a thorough cost analysis has emphasized the importance of comprehending the breakdown of expenses. Certain areas have been identified as presenting significant opportunities for cost-cutting measures. The significance of revenue segmentation in revenue analysis is highlighted, as it helps clubs identify important streams of revenue that can be leveraged to optimize their revenue plans. Our research on profitability provides a thorough understanding of a club's financial well-being, while benchmarking analysis helps determine their competitive position in the market. Financial scenarios are carefully examined through the use of cluster and time series models. Clustering models categorize institutions based on their revenue sources and offer valuable insights into different revenue models. Time series models allow us to create future sales projections by analyzing sales trends over time.

Based upon the dataset there is a diverse range of budgets and sponsorship types. In order to get more insight and make it more optimized, thorough research on the market is needed, to understand the market share, fan demographics to maximize the profit and retain profitable sponsors. Also, there should be consistent monitoring and analysis to check the performance of current sponsors and based upon the output further analyses can be done. A proper budget of sponsors should be set to start an initial companionship and split the budget between various sectors of the team.

The analysis resulted in various outcomes that can be looked upon to improve the performance of the squad as a whole. The data collected and analyzed was for the last 20 years and the pattern on points and the half-time scores and the end result were studied over the years. Exploratory data analysis was done to understand the team dynamics and study the goals scored, conceded and points won in the previous seasons. The predictive model that was built to find the probability of the win on the basis of halftime results and the opposition and also the venue being played on shows that it can be used to motivate the team to score more goals and go for a win. In order to increase the model accuracy, we can run various other models and increase the number of input parameters and take into consideration the time between the games, whether they are played on the weekend or the weekdays and how they impact the result. Overall, the exploratory analysis that was done helped to understand the past results and the team performed within various matrices.

Also, a linear regression model that attempts to predict the Total Goals scored by soccer players based on their Weight, Height, and Age. It was found that it is a significant predictor. So, this could help in player selection.

6. Conclusions and implications recommendations

The final deductions and potential ramifications of the given information will be discussed in this section, along with suggested courses of action. Based on the comprehensive study of the preceding sections, we have arrived at several recommendations for financial and team management, which we will outline in the following findings.

Cost Control: Regularly reviewing important expense items is crucial for cost control. In order to cut costs, clubs have the option of implementing cost-cutting measures, such as negotiating better contracts, finding sponsors for facilities and equipment, and reducing match and travel department expenses. Experiments 1 and 2 have demonstrated the effectiveness of these techniques in lowering costs without negatively impacting the team or overall functionality.

Diversification of revenue: In order for football clubs to remain financially stable, they must broaden their revenue streams beyond the typical sources such as ticket sales and broadcasting rights. Revenue diversification is key, and this can be achieved through various means, such as expanding marketing efforts, merchandise sales, and utilizing internet channels, as seen in Experiments 3 and 4.

Increase profitability: Optimizing operations is crucial for football teams to augment profits. To achieve this, Experiments 5 and 6 suggest that implementing more efficient training programs, incorporating new technologies, or modifying player transfer policies could be potential solutions.

Data-driven financial decision-making: The implementation of a recommender system can greatly aid football clubs in making well-informed financial decisions based on data analysis. This technology should be fully utilized to expedite and improve the accuracy of decision-making, particularly when it comes to player transfers and scouting. The efficacy of this system is demonstrated in Experiments 7, 8, and 9.

Annual Budget: The annual budget of the sponsors should align with the team's desired objectives. First understand the fan demographics and utilize the location to maximize the revenue. Understanding the correlation between higher budget and higher market share. It's not always correct for an organization having a higher budget will have a monopoly in the market.

It could be seen that Liverpool's defence and attack both lack the vagility and stability that it produced in the previous seasons. So, if Liverpool's football team is struggling with weak defence and a lack of attacking power, there are several strategies they could consider for improving their performance. Adjust the formation could be the one, Liverpool could consider adjusting their formation to add more defensive stability or attacking power. For example, they could switch to a more defensive-minded 5-3-2 or 4-2-3-1 formation to bolster their defence, or a more attacking-minded 4-3-3 or 3-4-3 formation to generate more chances to go forward, with additional firepower in the squad.

Sign new players: Liverpool could look to sign new players in the transfer window to address their weaknesses. They could target a new centre-back or full-back to improve their defence, or a striker or attacking midfielder to provide more firepower in the final third.

Focus on team shape: Liverpool could focus on improving their overall team shape and positioning to ensure that they are more solid defensively and more dangerous going forward. This could involve working on their defensive transitions and pressing, as well as their attacking movements and combination play.

Utilize set pieces: Liverpool could focus on improving their set-piece play to generate more chances and goals. This could involve developing new set-piece routines or practicing existing ones to make them more effective.

Work on mental toughness: Finally, Liverpool could work on improving their mental toughness and resilience, particularly in tight matches. This could involve working on their concentration, focus, and confidence to ensure that they are able to see out games and grind out results when needed.

First 45 min and conversion rate: Encourage the team to take a lead at halftime, as seen from the data that the results remain the same in the majority of cases. The team needs to improve the conversion rate from set pieces, hiring a dedicated set piece coach can be done. Also, the team needs to improve the away record if we have to challenge for the title.

In order for football clubs to not only survive but also succeed in the fiercely competitive sports industry, it is crucial for them to find a balance between their financial stability and team performance. By following the provided guidance, clubs can make better-informed decisions, ultimately leading to a more prosperous and flourishing organization.

ANNEXURE

Codes:

```
# define expense categories
categories = ['Excess Transfers Back', 'Other Expenses', 'Medical',
              'Competition Guarantees', 'Recruiting', 'Game Expenses and Travel',
              'Facilities and Equipment', 'Coaches Compensation',
              'Support and Admin Compensation w/Severance', 'Athletic Student Aid'

# calculate total expenses and percentages by category
totals = revenue_df[categories].sum()
percentages = totals / totals.sum()

# plot expense breakdown
plt.pie(percentages, labels=categories, autopct='%1.1f%%')
plt.title('Athletic Department Expenses Breakdown')
plt.show()

# print expense breakdown by category
print('Expense breakdown by category:')
for category in categories:
    print(f'{category}: ${totals[category]:,.0f} ({percentages[category]*100:.1f}%
```

```
# group the data by NCAA Subdivision and FBS Conference
grouped_data = revenue_df.groupby(['NCAA Subdivision', 'FBS Conference'])['Total E:

# create a pivot table with NCAA Subdivision as rows, FBS Conference as columns, ai
pivot_table = grouped_data.pivot(index='NCAA Subdivision', columns='FBS Conference

# plot a heatmap of the pivot table
sns.heatmap(pivot_table, cmap='YlGnBu')
```

```
[ ]  # define revenue categories
     categories = ['Other Revenue', 'Corporate Sponsorship, Advertising, Licensing',
                   'Donor Contributions', 'Competition Guarantees',
                   'NCAA/Conference Distributions, Media Rights, and Post-Season Football',
                   'Ticket Sales', 'Institutional/Government Support', 'Student Fees']

     # calculate total revenues and percentages by category
     totals = revenue_df[categories].sum()
     percentages = totals / totals.sum()

     # plot revenue breakdown
     plt.pie(percentages, labels=categories, autopct='%1.1f%%')
     plt.title('Athletic Department Revenue Breakdown')
     plt.show()

     # print revenue breakdown by category
     print('Revenue breakdown by category:')
     for category in categories:
         print(f'{category}: ${totals[category]:,.0f} ({percentages[category]*100:.1f}%)')
```

```
[ ] # define revenue categories
    categories = ['Other Revenue', 'Corporate Sponsorship, Advertising, Licensing',
                  'Donor Contributions', 'Competition Guarantees',
                  'NCAA/Conference Distributions, Media Rights, and Post-Season Football',
                  'Ticket Sales', 'Institutional/Government Support', 'Student Fees']

    # calculate total revenues and percentages by category
    totals = revenue_df[categories].sum()
    percentages = totals / totals.sum()

    # create a scatter plot
    fig, ax = plt.subplots()
    ax.scatter(categories, percentages)
    ax.set_xlabel('Revenue Category')
    ax.set_ylabel('Percentage of Total Revenue')
    ax.set_title('Revenue Breakdown by Category')
    ax.set_xticklabels(categories, rotation=90) # rotate x-axis tick labels by 90 degrees
```

```
[ ] # define revenue and expense categories
    revenue_categories = ['Other Revenue', 'Corporate Sponsorship, Advertising, Licensing',
                          'Donor Contributions', 'Competition Guarantees',
                          'NCAA/Conference Distributions, Media Rights, and Post-Season Football',
                          'Ticket Sales', 'Institutional/Government Support', 'Student Fees']
    expense_categories = ['Excess Transfers Back', 'Other Expenses', 'Medical', 'Competition Guarantees',
                          'Recruiting', 'Game Expenses and Travel', 'Facilities and Equipment', 'Coaches Compensation',
                          'Support and Admin Compensation w/Severance', 'Athletic Student Aid', 'Total Academic Spending (University-Wide)',
                          'Total Football Spending', 'Total Football Coaching Salaries', 'Athletics Related Debt',
                          'Annual Debt Service, Leases and Rental Fees on Athletic Facilities']

    # calculate total revenues, expenses, and profits
    revenue_df['Total Revenues'] = revenue_df[revenue_categories].sum(axis=1)
    revenue_df['Total Expenses'] = revenue_df[expense_categories].sum(axis=1)
    revenue_df['Profit'] = revenue_df['Total Revenues'] - revenue_df['Total Expenses']

    # plot profit and loss over time
    plt.plot(revenue_df['Year'], revenue_df['Profit'])
    plt.xlabel('Year')
    plt.ylabel('Profit ($)')
    plt.title('Athletic Department Profit and Loss Over Time')
    plt.show()

    # plot profitability by school
    df_sorted = revenue_df.sort_values(by='Profit', ascending=False)
    plt.bar(df_sorted['FBS Conference'], df_sorted['Profit'])
    plt.xticks(rotation=90)
    plt.xlabel('School')
    plt.ylabel('Profit ($)')
    plt.title('Athletic Department Profitability by School')
    plt.show()

    # cluster schools based on revenue and expense categories
    X = revenue_df[revenue_categories + expense_categories]
    scaler = StandardScaler()
    X_scaled = scaler.fit_transform(X)
    kmeans = KMeans(n_clusters=3)
    kmeans.fit(X_scaled)
    revenue_df['Cluster'] = kmeans.labels_
    fig, ax = plt.subplots()
    scatter = ax.scatter(revenue_df['Total Revenues'], revenue_df['Total Expenses'], c=revenue_df['Cluster'])
    plt.xlabel('Total Revenues')
    plt.ylabel('Total Expenses')
    plt.title('Revenue vs. Expense Clustering')
    legend = ax.legend(*scatter.legend_elements(),
                       loc="upper right", title="Clusters")
    ax.add_artist(legend)
    plt.show()
```

```
# define revenue and expense categories
revenue_categories = ['Other Revenue', 'Corporate Sponsorship, Advertising, Licensing',
                      'Donor Contributions', 'Competition Guarantees',
                      'NCAA/Conference Distributions, Media Rights, and Post-Season Football',
                      'Ticket Sales', 'Institutional/Government Support', 'Student Fees']
expense_categories = ['Excess Transfers Back', 'Other Expenses', 'Medical', 'Competition Guarantees',
                      'Recruiting', 'Game Expenses and Travel', 'Facilities and Equipment', 'Coaches Compensation',
                      'Support and Admin Compensation w/Severance', 'Athletic Student Aid', 'Total Academic Spending (University-Wide)',
                      'Total Football Spending', 'Total Football Coaching Salaries', 'Athletics Related Debt',
                      'Annual Debt Service, Leases and Rental Fees on Athletic Facilities']

# calculate total revenues, expenses, and profits
revenue_df['Total Revenues'] = revenue_df[revenue_categories].sum(axis=1)
revenue_df['Total Expenses'] = revenue_df[expense_categories].sum(axis=1)
revenue_df['Profit'] = revenue_df['Total Revenues'] - revenue_df['Total Expenses']

# create a new dataframe for conference data
conference_data = pd.DataFrame()

# calculate the average revenue and expense categories for the conference
for category in revenue_categories + expense_categories:
    conference_data[category] = [revenue_df[category].mean()]

# set the index of the dataframe to 'Conference'
conference_data.index = ['FBS Conference']

# reset the index of revenue_df
revenue_df = revenue_df.reset_index()

# append the conference data to our existing data
df = pd.concat([revenue_df, conference_data], axis=0)

# plot our data compared to the conference average
fig, ax = plt.subplots(figsize=(10,6))

for i, category in enumerate(revenue_categories + expense_categories):
    ax.bar(i, df.loc[df['School']=='Our School', category], color='b', alpha=0.5)
    ax.bar(i, df.loc['FBS Conference', category], color='r', alpha=0.5)

ax.set_xticks(range(len(revenue_categories + expense_categories)))
ax.set_xticklabels(revenue_categories + expense_categories, rotation=90)
ax.set_xlabel('Category')
ax.set_ylabel('Amount ($)')
ax.set_title('Comparison of Our School vs. Conference Average Revenue and Expense Categories')

plt.legend(['Our School', 'Conference Average'], loc='upper right')

plt.show()
```

```
[ ] # define revenue categories
    categories = ['Other Revenue', 'Corporate Sponsorship, Advertising, Licensing',
                  'Donor Contributions', 'Competition Guarantees',
                  'NCAA/Conference Distributions, Media Rights, and Post-Season Football',
                  'Ticket Sales', 'Institutional/Government Support', 'Student Fees']

    # filter the data to include only the revenue categories
    cluster_data = revenue_df[categories]

    # one-hot encode the categorical variable
    cluster_data_encoded = pd.get_dummies(cluster_data)

    # standardize the data
    scaler = StandardScaler()
    cluster_data_std = scaler.fit_transform(cluster_data_encoded)

    # apply PCA to reduce dimensionality
    pca = PCA(n_components=2)
    cluster_data_pca = pca.fit_transform(cluster_data_std)

    # perform clustering using KMeans
    kmeans = KMeans(n_clusters=3)
    kmeans.fit(cluster_data_pca)

    # visualize the clustering results
    plt.scatter(cluster_data_pca[:,0], cluster_data_pca[:,1], c=kmeans.labels_)
    plt.title('Revenue Clustering')
    plt.xlabel('PCA Component 1')
    plt.ylabel('PCA Component 2')
    plt.show()
```

```
[ ] # load the dataset
    df = pd.read_excel('NCAA Profit and Losses.xlsx', index_col='Year', parse_dates=True)

    # create a time series plot
    plt.plot(df.index, df['Total Revenues'])
    plt.xlabel('Year')
    plt.ylabel('Total Revenue')
    plt.title('Total Revenue over Time')
    plt.show()

    # decompose the time series
    decomposition = seasonal_decompose(df['Total Revenues'], model='additive', period=1)
    trend = decomposition.trend
    seasonal = decomposition.seasonal
    residual = decomposition.resid

    # plot the decomposed time series
    plt.subplot(411)
    plt.plot(df['Total Revenues'], label='Original')
    plt.legend(loc='upper left')
    plt.subplot(412)
    plt.plot(trend, label='Trend')
    plt.legend(loc='upper left')
    plt.subplot(413)
    plt.plot(seasonal, label='Seasonality')
    plt.legend(loc='upper left')
    plt.subplot(414)
    plt.plot(residual, label='Residuals')
    plt.legend(loc='upper left')
    plt.tight_layout()
    plt.show()

    # plot the autocorrelation and partial autocorrelation functions
    plot_acf(df['Total Revenues'], lags=20)
    plot_pacf(df['Total Revenues'], lags=20)
    plt.show()

    # fit a SARIMA model
    model = SARIMAX(df['Total Revenues'], order=(1,0,1), seasonal_order=(0,0,0,4))
    results = model.fit()

    # generate predictions
    predictions = results.predict(start=len(df), end=len(df)+4, dynamic=True)

    # plot the actual vs. predicted values
    plt.plot(df.index, df['Total Revenues'], label='Actual')
    plt.plot(predictions.index, predictions, label='Predicted')
    plt.legend(loc='upper left')
    plt.xlabel('Year')
    plt.ylabel('Total Revenue')
    plt.title('Actual vs. Predicted Total Revenue')
    plt.show()
```

REFERENCES

[1] Buchheit, M. (2014). The Numbers Game: Where Data and Football Analytics Meet. In M. Hughes, R. L. Franks, & M. Tenga (Eds.), Notational Analysis of Sport IV (pp. 109-116). Routledge.

[2] Decroos, T., et al. (2019). Actions Speak Louder Than Goals: Valuing Player Actions in Soccer. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 1851-1861).

[3] Fernandez-Navarro, J., et al. (2019). An Expected Possession Value Framework to Model Football Game Events. PLOS ONE, 14(10), e0223945.

[4] Duch, J., et al. (2019). Sports Analytics and Data Science: Winning the Game with Methods and Models. CRC Press.

[5] Koehn, S., et al. (2019). The Expected Influence Model: Quantifying Relative Performance in Soccer. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 1896-1906)

[6] Holman, V. What is Sports Analytics? *Agile Sports Analytics.* [Online] November 15, 2018. https://www.agilesportsanalytics.com/what-is-sports-analytics/.

[7] Dixon, M.J. and Coles, S.G. *Modelling Association Football Scores and Inefficiencies in the Football Betting Market.* 1997.

[8] Lago-Peñas, C. - Lago-Ballesteros, J. - Rey, E. *Differences in performance indicators between winning and losing teams in the UEFA Champions League*, 2011, Journal of Human Kinetics, Vol. 27, pp. 135-146.

[9] Harrop, K. and Nevill, A. *Performance indicators that predict success in an English professional League One soccer team*, 2014, Int'l Journal of Performance Analysis in Sport, Vol. 14, pp. 907-920.

[10] Mao, L. - Peng, Z. - Liu, H. - Gómez, M.-A. *Identifying keys to win in the Chinese professional soccer league.* 2016, Vol. 16, pp. 935-947.

[11] Tax, N. and Joustra, Y. *Predicting the Dutch Football Competition Using Public Data: A Machine Learning Approach.* 10, 2015, Transactions of knowledge and data engineering, Vol. 10.

[12] McCabe, A. and Trevathan, J. *Artificial Intelligence in Sports Prediction* Information Technology: New Generations, 2008. pp. 1194-1197.

[13] Hucaljuk, J. and Rakipovic, A. *Predicting football scores using machine learning techniques.* 2011. MIPRO, 2011 Proc. 34th Int'l Convention.

[14] Goddard, J. *Regression models for forecasting goals and match results in association football*, Elsevier B.V., 2005, Int'l Journal of Forecasting, Vol. 21, pp. 331-340.

[15] Eggels, H. - van Elk, R. - Pechenizkiy, M. *Explaining soccer match outcomes with goal scoring opportunities predictive analytics*, 2016.

[16] Lepschy, H. - Wasche, H. - Woll, A. *How to be Successful in Football: A Systematic Review.*1, 2018, The Open Sports Sciences Journal, Vol. 11, pp. 3-23.

[17] Bekris, E. - Gioldasis, A. - Gissis, I. - Komsis, S. - Alipasali, F. *Winners and losers in top level soccer. How do they differ?* 2014, Journal of Physical Education and Sport, Vol. 14, pp. 398-405.

[18] Hvattum, L.M. and Arntzen, H. *Using ELO ratings for match result prediction in association football.* 2010, Int'l Journal of Forecasting, Vol. 26, pp. 460-470.

[19] Constantinou, A. and Fenton, N. *Towards smart data: Improving predictive accuracy in long-term football team performance.* 2017, Knowledge-Based Systems, Vol. 124, pp. 93-104.

[20] Van Haaren, J. and Davis, J. *Predicting the Final League Tables of Domestic Football Leagues.* 2015. 5th int'l conf. mathematics in sport. pp. 202-207.

[21] Oberstone, J. *Differentiating the Top English Premier League Football Clubs from the Rest of the Pack: Identifying the Keys to Success.* 2009, Journal of Quantitative Analysis in Sports, Vol. 5.

[22] Kringstad, M. and Olsen, T.-E. *Can sporting success in Norwegian football be predicted from budgeted revenues?*